(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 October 2002 (17.10.2002)

PCT

(10) International Publication Number
WO 02/082271 A1

(51) International Patent Classification[7]: G06F 11/30

(21) International Application Number: PCT/US02/10615

(22) International Filing Date: 3 April 2002 (03.04.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/281,881 5 April 2001 (05.04.2001) US

(71) Applicant: AUDIBLE MAGIC CORPORATION [US/US]; 985 University Avenue, Suite 35, Los Gatos, CA 95032 (US).

(72) Inventor: SCHMELZER, Richard, A.; 1080 Juniper Avenue, Boulder, CO 80304 (US).

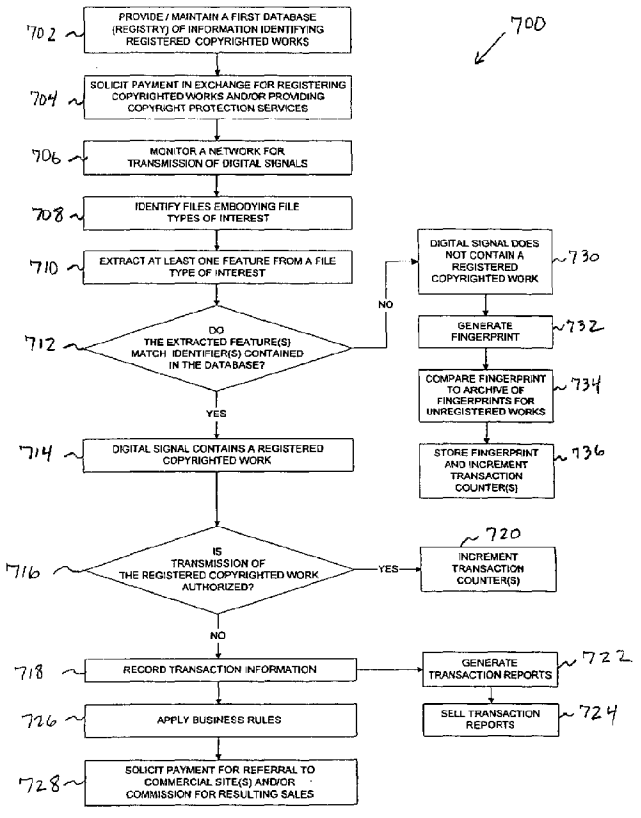
(74) Agent: SIERRA PATENT GROUP, Ltd.; P.O. Box 6149, Stateline, NV 89449 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: COPYRIGHT DETECTION AND PROTECTION SYSTEM AND METHOD



(57) Abstract: A method for detecting against unauthorized transmission of digital works comprises the steps of maintaining a registry of information permitting identification of digital copyrighted works (702), monitoring a network (706) for transmission of at least one packet-based digital signal, extracting (710) at least one feature from the at least one digital signal,comparing (712) the extracted at least one feature with registry information and applying business rules (726) based on the comparison result.

WO 02/082271 A1

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

COPYRIGHT DETECTION AND PROTECTION SYSTEM AND METHOD

PRIORITY CLAIM

This application claims the benefit of United States Provisional Application Serial No. 60/281,881, filed April 5, 2001, and United States Patent Application filed April 3, 2002, Serial No. not yet available.

BACKGROUND

1. Field

The field of the present invention relates to processing digital data. More particularly, the field of the present invention relates to identifying, reporting and/or protecting digital works from unauthorized transmission and/or copying, such as over networks or network segments connected to the Internet.

2. Background

Technological developments such as peer to peer file sharing have revolutionized the exchange of information over digital networks such as the Internet. The result has been a virtual explosion of copyright violations, as intellectual property is transmitted to individuals not licensed to receive and use it. Once copyrighted content is available on the Internet, that content is exposed to piracy. The unlicensed use of digital copyrighted works online is a growing, major concern to the owners of these properties. Current peer-to-peer file sharing technology facilitates widespread copyright infringement of various works including songs, images, and movies. At the same time, security measures placed into widespread use have been defeated. For example, DVD encryption was "hacked" by mid-2000, resulting in the digital copying and distribution of movies without regard for copyright. As a result of the widespread availability of digital works on computer networks, artists and companies affiliated with them receive no payment for distribution of copyrighted works on an unprecedented scale.

In response to the growing copyright infringement problem tied to unregulated peer-to-peer file sharing, copyright owners have sought and obtained legal relief, including injunctive relief, against peer-to-peer facilitators such as Napster. Some copyright owners have further requested that network operators, including colleges and universities, block

affiliated with them receive no payment for distribution of copyrighted works on an unprecedented scale.

In response to the growing copyright infringement problem tied to unregulated peer-to-peer file sharing, copyright owners have sought and obtained legal relief, including injunctive relief, against peer-to-peer facilitators such as Napster. Some copyright owners have further requested that network operators, including colleges and universities, block access to peer-to-peer sites to prevent further copyright infringement. At the same time, however, there exist substantial non-infringing uses for peer-to-peer file sharing, including exchange of creative works that exist in the public domain (such as may exist through expiration or abandonment of copyrights, for example) and/or uses that have been expressly permitted. If aggrieved copyright owners prevail in their legal battles against peer-to-peer facilitators, then such facilitators may be forced to stop operating irrespective of the content they provide.

The injunction entered against Napster in March 2000 by a federal judge in San Francisco, California has ordered the company to remove copyrighted sound recordings from its system. The recording industry has been given the duty to provide lists containing the titles, names of artists, file names, and ownership rights of recordings, and Napster, shortly after receiving such identification, is responsible for blocking those materials from its system. Yet compliance with this name-based regime has already proven difficult, since there exists no file-naming standard and file names can be easily manipulated with known method presently in use. The inclusion of metadata (data about data, usually constituting text embedded in an audio file or stream to represent information such as artist name, album name, track name, etc.) in selected audio works may aid in identifying works even if file names are changed. However, metadata is only present on newer works, and essentially amounts to a more sophisticated extension of file naming technology that is subject to manipulation and hacking.

A potential alternative to relying on file naming technology for identifying digital works on computer networks is an identification technology known as watermarking. A watermark is digital information that is embedded into a file in such a way that it does not affect human perception of the content but is easily detectable by machines. One advantage offered by watermarking is its easy recognition. However, drawbacks of watermarking technology include its inability to protect the huge amount of previously released audio content, and its susceptibility to hacking. Once a watermark is disabled or removed from a creative work by a hacker, the resulting product is unprotected.

A different identification technology known as content-based identification ("CBID"), relying on the content of creative works, represents yet another alternative to file naming technology. For example, when applied to audio works, CBID analyzes acoustic qualities. Various CBID techniques may be used to characterize the qualities of sound perceived by a listener. A typical approach is to analyze the spectrum of a sound, such as by measuring the loudness of each frequency contained in a multi-frequency sound.

A more compact CBID technology involves creation of a "fingerprint" from a creative work that is compact from a data perspective, yet preserves distinguishing characteristics that may be used to positively identify a unique audio file. Many simple fingerprinting methods have been developed, such as spectral averaging, for example. In using these simpler methods, however, a substantial amount of information about the audio work is lost. Great care must be taken in applying a particular CBID method for a number of reasons: not only to ensure only accurate identification, but also to ensure that compressed versions of an audio file can be identified, and to avoid known evasion techniques such as adding a small segment to the beginning of an audio file. A more sophisticated CBID technology would be appropriate to address these concerns.

One structural application of a sophisticated CBID fingerprinting method for audio data is found in U.S. Patent No. 5,918,223, issued to Blum et al., the disclosure of which is hereby incorporated by reference as if fully set forth herein. The patent provides a system

and method for performing analysis and comparison of audio data files based upon the content of the data files. However, U.S. Patent No. 5,918,223 by itself does not address a comprehensive solution to regulating distribution of digital copyrighted works. Moreover, U.S. Patent No. 5,918,223 expressly relates to audio information, and does not address the similar but distinct problems with regulating online distribution of copyrighted works such as motion pictures, still images, games, software, and other media.

Regarding movies, the transformation taking place in the motion picture industry from VHS video to digital DVD format has led to the spread of illegally shared copies of movies online. While a universal DVD encryption system has been adopted by the motion picture industry to block the online trade of illegal DVD content, as mentioned previously, decryption software such as De-Content Scrambling System (DeCSS) is readily available online. Moreover, technologies such as DivX allows users to take the decoded movie and copy the material onto a CD-ROM for home use through a standard Internet connection. The Motion Picture Association of America (MPAA) has moved aggressively to stop the illicit trade of movies online. The MPAA has sued online sites and chat rooms that offer pirated movies, as well as sites offering shared movie files, under the recently adopted Digital Millennium Copyright Act.

With regard to images, photo communities are quickly becoming a favorite new tool of online users, as such communities allow users to post, print, and share their photos online with other subscribers. The explosive growth in digital camera use has greatly expanded the popularity of these photo communities. While many sites promote their usefulness in sharing family moments and other important events online, some estimates provide that, in reality, half of all images posted on these sites are copyright-protected images, and are being posted, printed and shared illegally.

In summary, peer-to-peer file sharing technology offers unprecedented ease in exchanging information over digital networks. Unfortunately, this technology also permits intellectual property rights to be infringed on a widespread scale. Without a comprehensive

protection system in place to prevent further infringement of intellectual property rights, if intellectual property owners prevail in their ongoing legal battles against peer-to-peer providers, then the benefits of peer-to-peer file sharing may be lost to everyone. In light of all of the considerations discussed above, it would be desirable to provide a reliable and secure system for enabling intellectual property owners to distribute digital materials while preventing infringement of intellectual property rights. Preferably, such a system would permit intellectual property owners to choose whether distribution of particular works should be unrestricted, restricted, or disallowed entirely.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a high-level schematic of a copyright protection system according to a first embodiment.

FIG. 2 is a sample report generated by a copyright protection system, the report including several data fields useful to record a transmission transaction.

FIG. 3 is component architecture schematic for a portion of a copyright protection system directed to monitoring a multi-session digital signal.

FIG. 4 is a schematic of a copyright protection system including implementation details for content type recognition and identification, in accordance with a second embodiment.

FIG. 5 is a schematic of a copyright protection system according to a third embodiment.

FIG. 6 is a process flow diagram for a hierarchical method useful with a copyright protection system to assess whether a digital file contains a registered copyrighted work.

FIG. 7 is a process flow diagram for obtaining and entering information useful to a copyright protection system into a database.

FIG. 8 is a schematic of a copyright protection system having a distributed architecture for monitoring multiple watched networks.

FIG. 9 is a process flow diagram for a method of conducting a business enterprise through the provision of copyright protection services or a copyright protection system.

FIG. 10 is a generalized data flow diagram for use with a Stochastic Audio Matching Mechanism.

FIG. 11 is a process flow diagram for extracting feature vectors comprising Mel Frequency Cepstral Coefficients.

FIG. 12a is a first portion of an annotated sequence diagram for extracting features from a digital audio work according to a Stochastic Audio Matching Mechanism.

FIG. 12b is a second portion of the annotated sequence diagram of FIG. 12a.

FIG. 13a is a graph plotting frequency versus time for a first musical piece performed by a first artist.

FIG. 13b is a graph plotting frequency versus time for a second musical piece performed by a second artist.

FIG. 14 is an annotated sequence diagram for generating a model from a digital audio work according to a Stochastic Audio Matching Mechanism.

FIG. 15 is an annotated sequence diagram for identifying a digital audio work according to a Stochastic Audio Matching Mechanism.

FIGS. 16-21 illustrate examples of screenshots that may be viewed by an intended recipient of unauthorized content in the context of a peer-to-peer file-sharing network.

DETAILED DESCRIPTION

FIG. 1 generally illustrates a copyright protection system ("CPS") 100 according to a first embodiment for monitoring a network segment 102 bearing at least one packet-based digital signal in accordance with one aspect of the CPS 100. In other aspects of the CPS 100, the monitoring point for a data transaction may be at points other than a network segment of a communication. For example, the monitoring point of the CPS may be a server on a community website that monitors the uploads of audio, image, video or other digital

content. The same community website may alternatively monitor downloads of such data. Alternatively, the monitoring point may be a peer or client computer in a peer-to-peer file sharing network. In yet another embodiment, the CPS 100 may be integrated or linked with a search engine such as Excite® or Infoseek® that monitors search requests and performs one or more of the actions of monitoring, recording or blocking based on the nature of the request and the likelihood that it involves transacting copyright protected material. The network segment 102 is routed through a network appliance 104 that monitors digital signals borne by the segment 102. While FIG. 1 suggests that the network appliance 104 receives in-stream communications from the network segment 102, in other embodiments the network appliance 104 may alternatively receive mirrored data from a network. For an in-stream configuration such as is suggested by FIG. 1, each network appliance 104 would typically communicate with the network segment 102 through a router (not shown) having content recognition capability, such as routers commercially available from companies such as Cisco Systems or Alteon WebSystems (product information available at http://www.cisco.com and http://www.alteonwebsystems.com, respectively). Preferably, any digital signals borne by the network segment 102 are periodically sampled to obtain a frame of sample data on each occasion. As noted in U.S. Patent 5,918,223, various window periods may be used for each frame, but each frame advantageously contains several milliseconds of data. A sampled frame is provided to a content recognizer 116, preferably part of the network appliance 104 that recognizes defined content types. Exemplary content types include .mp3, .avi, .asf, .ogg, but searching and recognition of practically any recognizable file type bearing audio, video, or image data, or digital text, or software, may be addressed by the content recognizer 116.

Upon recognition of the appropriate file type, a sampled frame is then provided to a media analysis system 126. The purpose of the media analysis system 126 is to assess the content of a digital file. While content may be determined according to different methods, one desirable method is to use digital content-based fingerprinting if sufficient processing

resources are available. Preferably, a fingerprint is generated for the frame by the media analysis system 126 to aid in identifying the content of the frame. A generated fingerprint may then be compared with an archive of fingerprints for registered copyrighted works. "Registered copyrighted works" as used herein refers to digital works registered with or by a CPS provider or service provider. The existence of a fingerprint archive suggests that, in a preferred embodiment, copyrighted works should be registered with the provider of the CPS 100, and reference fingerprints should be generated from registered copyrighted works, before seeking to detect the transmission of particular works in a network segment 102. If the comparison between the fingerprint of the frame and an archived fingerprint yields a match, thus signifying the transmission of a registered copyrighted work along the network segment 102, then transmission information is recorded in a content transmission recording device 110.

As illustrated in the sample report provided in FIG. 2, several data fields identifying a transmission transaction may be recorded, including, for example, any one or more of the following:

a) Source IP Address: the Internet Protocol (IP) address from which the recognized content was transmitted;

b) Destination IP Address: the IP address to which the recognized content was transmitted;

c) Date Transmitted: the date the recognized media was transmitted;

d) Time Transmitted: the time the recognized media was transmitted;

e) Content / Media Name: The name or title of the content whether audio, video, still image, or other type;

f) Artist Name: The name of the artist (when appropriate) if the work is a copyrighted work already registered with the CPS provider;

g) Album Name: The name of an album (if appropriate) associated with a registered copyrighted (e.g., audio) work;

h) Record Label: The name of an album (if appropriate) associated with a registered copyrighted (e.g., audio) work;

i) Various Meta-Data: Distributor name, producer name, studio name, etc., such as may be found attached to a .id3 or .md5 file or tag associated with the copyrighted work;

j) Unauthorized Count: The number of unauthorized downloads organized in various ways, such as by day, week, month, location, IP address, etc.;

k) Redirected Count: The number of redirected attempted downloads organized in various ways, such as by day, week, month, location, IP address, etc.

Referring back to FIG. 1, various components of the CPS 100 may be optionally located remotely to one another and connected by a network connection 107. For example, certain components such as the network appliance 104 and a content recognizer 116 may be located at a first monitored network data center 121, while the remaining components 126, 146 may be located within a separate administrative network data center 123. FIG. 3 illustrates a preferred embodiment of a component architecture for a portion 280 of a CPS 100, such as the CPS 100 depicted in FIG. 1, the illustrated portion 280 being useful for monitoring a multi-session signal such as may be transmitted along a high bandwidth network segment. A high bandwidth network connection 262, preferably anticipated to operate at a convenient, commercially available speed, preferably greater than 28 kbps, communicates at least one packet-based digital signal to a first statefull session-binding load balancer 264 that separates the greater network stream into individual TCP or UDP sessions and binds those sessions to a specific processing unit (e.g., 268, 269, or 270) in the next layer. Connections 265, 266, 267 communicate individual network sessions to content-type recognition and identification servers 268, 269, 270, each having at least one processor. Each server 268, 269, 270, which preferably includes at least one processor, executes content-type recognition and content identification services. Within the servers 268, 269, 270, the raw IP data packets are assembled (or re-assembled), the packets are analyzed for

presence of media types likely to contain copyrighted content using a content type recognition service, and the media content is identified using a content identifier service.

Though not shown in FIG. 3, the servers 268, 269, 270 preferably have further connections (remote or local) to a stored data repository to facilitate content comparison with known identifiers for copyrighted content using one or more processors. From the servers 268, 269, 270, packets may be communicated to a second statefull session-binding load balancer 274 that reassembles the various separated packets into a single network stream 275. Use of a second load balancer 274 to reassemble the separated packets into a single network stream 275 is generally only necessary if the portion 280 of the CPS 100 depicted in FIG. 3 is configured to operate in-stream. In such a case, the high bandwidth network connection 262 would typically be provided to the load balancer 264 by way of a router (not shown). Alternatively, if the CPS portion depicted in FIG. 3 receives mirrored network data, then the second load balancer 274 would be unnecessary, as there would be no need to reassemble separated packets into a single network stream 275 as the data is generally already streamed to its intended destination. Although not shown, additional redundant load balancers 264, 274, servers 268, 269, 270, and/or connections 265, 266, 267, 271, 272, 273 may be provided to provide failover (backup) capability in case one or more primary devices should fail.

FIG. 4 depicts a preferred embodiment of a detailed implementation of a CPS 100, 200, omitting (for the sake of simplicity) load balancing devices such as are shown in FIG. 3 to focus on a single session. An incoming network data stream 202 carrying at least one packet-based digital signal, preferably separated by session, is provided to a network appliance 204. The network appliance 204 may be characterized as a server, and the various operational blocks contained within the appliance 204 may be characterized as services, each amenable to at least partial performance in software routines. The network appliance 204 includes at least one processor that, in conjunction with memory, operates software code for performing various operations on the digital signal. The processor may comprise any type

of computer, and has processing characteristics dependent upon processing requirements for performing the various tasks discussed herein. It may comprise, e.g., a computer, such as a workstation including the type manufactured by Sun Microsystems, a main frame computer, or a personal computer such as the type manufactured by IBM® or Apple®.

The term "processor," as used herein, refers to a wide variety of computational devices or means including, for example, using multiple processors that perform different processing tasks or have the same tasks distributed between processors. The processor(s) may be general purpose CPUs or special purpose processors such as are often conventionally used in digital signal processing systems. Further, multiple processors may be implemented in a server-client or other network configuration, as a pipeline array of processors, etc. Some or all of the processing is alternatively implemented with hard-wired circuitry such as an application-specific integrated circuit (ASIC), a field programmable gate array (FPGA) or other logic device. In conjunction with the term "processor," the term "memory" refers to any storage medium that is accessible to a processor that meets the memory storage needs for a system or system component for performing the functions described herein. Preferably, the memory buffer is random access memory (RAM) that is directly accessed by the processor for ease in manipulating and processing selected portions of data. Preferably, the memory store comprises a hard disk or other non-volatile memory device or component.

The network appliance 204 may be installed either in series with or receiving mirrored data from a high bandwidth network segment. Preferably, a packet input receiver 206 accepts the input of a network data stream 202. Associated with the packet input receiver 204 is a TCP stream buffering / assembly service 206 that identifies the packet type of the input signal, and if the type is TCP, also provides storage buffering as needed and assembles the synchronized packet stream. Thereafter, a data extraction service 210 extracts the data from synchronized network packets, and then a data buffering service 212 assembles and buffers the data from the incoming packets.

Following data assembly and buffering, a content lookup service 214 communicates part or all of the data to a content type recognizer service 216. Any portion not communicated with the content type recognizer service 216 may be communicated instead to a packet output service or transmitter 250. The content type recognizer 216 preferably has multiple associated recognizers 218, 220, 222, 224, 225 to recognize file types of interest including, for example, .mp3, .avi, .asf, .ogg, and other types, respectively.

Following content type recognition, packets are forwarded to a remote or local content identifier service 226 preferably having multiple associated identifiers 228, 230, 232, 234, and 235 to identify content borne by file types of interest including, for example, .mp3, .avi, .asf, .ogg, and other types, respectively. Preferably, the content identifier service 226 is linked to a fingerprint generator service 240. While the fingerprint generator service 240 is illustrated as a distinct service from the content identifier 226, the two services optionally may advantageously be combined. Within the fingerprint generator 240, a content-based fingerprint comprising identifying features may be generated for a frame of data, and then forwarded to a content comparator 242. It may not be necessary to utilize a fingerprint generator 240 for identifying all digital files borne by the network data stream 202, as will be discussed hereinafter. Consequently, the content identifier 226 preferably includes a separate link to the content comparator 242 that is independent from the fingerprint generator 240.

The content comparator 242 is in communication with a database 244 of stored content identifiers, preferably by a high-speed network connection. The database 244 preferably includes database software such as is commercially available from Oracle® Corporation operating on one or more high-speed computers with expandable high-speed storage capability. The database 244 contains stored content-based identifiers, preferably including fingerprints, for copyrighted works registered with a CPS provider such as ipArchive™. For example, when a copyrighted song is registered with or by a CPS provider, the CPS provider would generate entries in the database 244 to assist in identifying the song,

preferably including at least one fingerprint from the song's content according to a CBID method, such as the method disclosed in U.S. Patent No. 5,918,223 issued to Blum et al. The CPS provider preferably indexes identifiers including fingerprints to registered works in the database 244. Fingerprints may be generated with a content identifier 226 with fingerprint generator 240, or with a media analysis system 326 such as provided in FIG. 5.

Returning to the content comparator 242, its function is to compare a content identifier (such as, for example, a fingerprint generated by the fingerprint generator 240) from the incoming data stream 202 and query the database 244 for stored identifiers for registered copyrighted works, and then determine whether the incoming data stream 202 matches with any archived content. If a match is found, then further actions may be necessary based on business rules associated with the identified content of, the data stream 202. Information identifying users, destination addresses, and/or passwords authorized to receive registered copyrighted content may be stored with the database 244, or, more preferably, in a separate database (not shown) communicating with the content comparator 242. The user/address/password information may be queried by the content comparator 242 to determine whether the data stream 202 is authorized.

A content transmission reporter 245 is preferably provided to record transmission information for copyright enforcement, record keeping, or other purposes. Information such as is listed above in connection with FIG. 2 may be stored, and reports such as the exemplary report provided in FIG. 2 may be generated. If the data stream 202 is not authorized, then one or more actions may be taken according to pre-defined business rules. Actions that might be taken according to pre-defined business rules, either separately or one or more in combination include, for example, recording, reporting and/or blocking a transmission, sending a generalized message to the source and/or recipient addresses involved with the unauthorized transaction, and sending a message informing a recipient address of (or redirecting a recipient address to) a commercial site where the desired copyrighted work may be purchased.

To facilitate messaging, a message generator 246 in communication with a packet output service or transmitter 250 is preferably provided. Preferably, messages are transmitted by way of an instant messaging protocol, such as the instant messenger associated with software distributed by www.napster.com, or AOL®. An alternative means for transmitting a message to a user is to send a message to a client application on the computer desktop of a user intended to receive the content, the client application including some communication capability. The CPS may detect an available client application, and then send the message accordingly. For example, the system may detect an Internet Explorer® on the user's desktop and send an HTML message to the user via the user's Internet Explorer®.

A transmitted message preferably provides instructions, or, more preferably, a link to a commercial site, for purchasing a license to the copyrighted work. In one embodiment, the recipient of the message is provided the option of contesting the blocking of the content. If the recipient chooses to contest the block, a return message is sent to the CPS 100, which then may immediately commence transmission of the digital data to the recipient. Alternatively, the CPS 100 may forward the contested data stream for additional identification processing or to an administrator of the CPS for review. In one preferred embodiment, the recipient is provided a small sample of both the transmitted content and the content to which it matched to enable the recipient to make an evaluation of whether to contest the block. For example, if the content is an image, thumbnails of the image and the matched image may be presented to the recipient side by side on the recipient's browser.

FIGS. 16-21 illustrate examples of screenshots that may be viewed by an intended recipient of unauthorized content in the context of a peer-to-peer file-sharing network. FIGS. 16 and 17 depict examples of screenshots as may be viewed by a user using a peer-to-peer file sharing client application (FIG. 16 for Napster and FIG. 17 for iMesh). The screenshots depict a list of songs that the intended recipient may choose to receive. In FIG. 18, a file sharing client application (e.g., such as for Napster) includes a window that depicts

the status of a file transfer. When the CPS intercedes in the transfer, the intended recipient may view a "Transfer error!" message on the client application. In one embodiment, this may be the complete and only message that is communicated b the CPS to the intended recipient. The intended recipient may not even be aware that the content has been affirmatively blocked, as the message may appear to indicate a communication problem or fault. Similarly, in FIG. 19, the message received is "Timed out!," which may or may not indicate to the content's intended recipient the reason for the failed transmission of the content. The specific action taken may depend on business rules associated with the content. The business rule may be construed to only report on the transmission and take no action to interfere with the transmission.

FIGS. 20 and 21 depicts examples of screenshots of windows that, in one alternative embodiment, may be presented to an intended recipient of unauthorized content. The windows preferably provide options to the viewer for obtaining the desired content from other sources, which are authorized to distribute the desired content, although typically for a fee.

Blocking or interrupting an unauthorized transmission may also be performed by way of the message generator 246, such as by transmitting a TCP/IP reset. This well-known technique is a form of IP spoofing in which the message generator 246 alternately gives the appearance that it is the transmitting device and the receiving device associated with a transaction, and then forges packets designed to terminate a TCP stream. According to this blocking method, an unlicensed and unauthorized destination address or recipient may be prevented from receiving full transmission of a specific registered copyrighted work. The forged packets are independent of any assembled content-bearing packets that may also be provided from the packet output service or transmitter 250 to a continued network traffic stream 260.

As noted previously, a continued network stream 260 suggests that the network appliance 204 is installed in-stream along a network segment. However, the appliance 204

may also be configured to receive mirrored network data, in which case the need to continue transmission of reassembled packets through the packet output service or transmitter 250 to a continued network stream 260 may be reduced or obviated. FIG. 5 is a schematic representation of an alternative copyright protection system 300. An incoming network stream 302 connected to the Internet 301 is routed to a media recognition system 316 provided at a network watchpoint. The media recognition system 316 includes an input receiver (not shown) for receiving an incoming network stream 302. If the media recognition system 316 is placed in-stream to capture all network communications, then an output transmitter (not shown) for transmitting the continued network stream 303 en route to a watched network 305 is preferably provided. The media recognition system 316 may also be configured to receive a mirrored network data stream according to conventional techniques. An in-stream approach requires additional, often expensive routing hardware (not shown), and may have a potential drawback of introducing latency into the monitored network stream. A potential benefit of an in-stream approach is that it may facilitate blocking of an entire transmission before any portion of it is transmitted to the watched network. The latter approach, implemented using mirrored network data, is preferred if it can be implemented at sufficient speed to render it effective at identifying and taking action against unauthorized transactions before such transactions are completed.

Preferably, multiple networks may be monitored by the copyright protection system 300 with additional media recognition systems 316 (such as embodied in the multiple network appliances 602, 604, 606, 608 shown in FIG. 8) each monitoring a segment of a different network but communicating with common analysis systems and/or a common transaction request broker. Each media recognition system 316 advantageously monitors a network 305 for traffic in digital files such as, for example, video, audio, image files and other digital content.

If a file type of interest is detected by the media recognition system 316, then any portion of the signal bearing such a file may be provided to the content analysis system 326

to perform content identification. There, separate media analysis subsystems 328, 330, 332 are provided for analyzing images, audio, and video or other media (including software) respectively. Image identification may be facilitated by use of the Ereo Exacta-Match system, developed by and commercially available from Ereo. Audio identification may be performed by application of the methods disclosed in U.S. Patent No. 5,918,223, issued to Blum et al. or alternatively with the Stochastic Audio Matching Mechanism (SAMM) discussed below. Video identification may be facilitated by applying one or both of the above-mentioned CBID methods to the audio portion of the video file, if any. Other digital works, such as digital text or software, may be identified by any number of methods as are known in the art.

The media analysis system 326 preferably includes a capability of generating CBID fingerprints for digital media, whether such media is obtained from an incoming network stream 302 by way of the media recognition system 316, or obtained from a raw media storage service 340. Preferably, the media analysis system 326 also includes storage capability to store content identifiers or fingerprints for registered copyrighted works, such as may be stored in and forwarded by the raw media storage service 340. The media storage service 340 preferably contains a raw media storage archive or database 338 and a raw media storage system manager 339 for managing transactions with the archive or database 338.

Returning to the media analysis system 326, a further function of the system 326 is to compare identifiers, preferably including fingerprints, extracted from the network stream 302 and from registered copyrighted works (such as are stored in the media storage service 340) to determine whether the network stream 302 contains any registered copyrighted content. If the media analysis system 326 finds a match in making this comparison, then it may forward information regarding the transaction to a transaction database service 344.

Within the transaction database service 344, a database 345 stores all media received by the media analysis system 326. The database 345 is preferably relational to facilitate

dimensional reporting, and preferably also permits high volume updates. A transaction recording and management service 343 is provided to manage queries to the database service 344 and also to manage data recordation in the database 345. Preferably, a data enrichment service 347 in communication with the database service 344 is provided to facilitate either automatic or manual addition of information potentially useful to the CPS (such as according to the method provided in FIG. 7).

A transaction reporting service 348, also is communication with the database service 344, is preferably provided to define and execute queries for generating reports including, for example, the transaction information provided in FIG. 2. Preferably, transaction reports may be sold by the CPS provider to owners of copyrighted works to communicate information useful for maximizing opportunities and revenue from the copyrighted works. An urgent or scheduled report forwarding service 349 is preferably provided and in communication with the transaction reporting service 348 to coordinate generation of urgent or scheduled reports. Preferably, reports may be delivered by way of email or another active, preferably electronic, delivery system to a client 352.

The transaction reporting service 348 is preferably in connection with a CPS transaction request broker service 350 that coordinates and manages various components of the CPS 300. The broker service 350 may be used to handle requests from the transaction reporting service 348, coordinate and/or manage operation of the media analysis system 326, handle requests of the transaction recording service 344, coordinate operations and data flows associated with the media storage service 340, and finally handle requests by and from the client 352. The client 352 preferably includes a web application interface providing access to intellectual property owners, reporting subscribers, and/or the community at large.

Reference has been made in the foregoing discussions to identifying the presence of a copyrighted work in a digital signal by way of content-based fingerprints. Such a methodology (as was described, for example, in connection with FIG. 1) provides but one way of performing content identification. While the method described in connection with

FIG. 1 is highly accurate, it may not be optimal to apply such a method to all digital files borne by a network segment due to the generally processor-intensive nature of fingerprint generation and comparison. If a copyright protection method is applied in-stream to intercept network traffic, then ensuring rapid identification speed is desirable to minimize latency.

Alternatively, if a copyright protection method is applied to mirrored network traffic, then it is important to ensure that content for a particular transaction in a registered copyrighted work is identified before the entire transaction is completed. For example, in the case of an unauthorized attempt to download a digital movie over a network, preferably the content of the movie is identified before the download is completed. Given limited processing resources, as traffic over a network increases, it may become difficult to generate and compare fingerprints for all network transactions with acceptable speed. Consequently, resort to a hierarchical method to assess the likely content of a digital signal being transmitted over a network may be desirable to ensure acceptable speed with finite processing resources.

FIG. 6 illustrates one embodiment of a hierarchical identity assessment method 400 that may be used in a CPS 100, 200, 300. A guiding principle of this method is to start with less processor-intensive steps to assess whether the monitored transmission contains a registered copyrighted work, and then to progress to more processor-intensive steps only if early steps do not indicate a match. Preferably, the method depicted in FIG. 6 is embedded in a software routing that may be operated on a computer processor, such as is contained in the network appliance 204 illustrated in FIG. 4. The method illustrated in FIG. 6 assumes that content type, file name, file size, IP addressing, any metadata, and/or watermarks may be discerned or extracted from a digital sample. Preferably, as a precursor to any assessment of the digital content that is transmitted, actions such as content blocking or content transmission reporting may be performed based on other aspects or attributes of the data stream. For example, an action may be taken based on the source IP address. Content

blocking, for example, may be performed based on protocol (e.g., Napster, Gnutella, etc.). Alternatively, content transmissions may be acted on based on the Internet Service Provider such as AOL®, used by the sender or the intended recipient of the content.

Utilizing file naming as one assessment criterion, the first step 402 is to compare the file name of the sample to file names of registered copyrighted works contained in a database (such as the database 244 illustrated in FIG. 4). If the file name of the digital sample matches a name in the database, then a checking comparison step 404 is preferably performed to compare the file size for the digital sample to the expected file size of the registered copyrighted work bearing that name in the database. If both the file name and file size appear to match, then the likelihood that the digital sample contains a registered copyrighted work considered is high, and a file match may be established according to block 422. Comparison of file names and file sizes is generally straightforward and does not consume substantial processing resources. Alternatively, the determination as to whether a match exists may be based only on the filename or the file size.

If the file name and file size do not both match, then a second assessment criterion involving a history of unauthorized transactions from a particular source address is preferably applied, according to step 406. As discussed previously, information recording various aspects of transactions in copyrighted data may be maintained in a database, such as the database 244 illustrated in FIG. 4. Representative aspects that may be recorded include the source and recipient IP addresses, the type and identity of copyrighted files, and the number and frequency of transactions or attempted transactions. If a particular source IP address generates a history of unauthorized transactions, especially involving files of a certain type, then the likelihood is elevated that a data stream emanating from that source IP address contains unauthorized copyrighted material. Accordingly, steps 406 and 408 examine whether a source IP address has a history of unauthorized transactions, and, if so, whether the file type and/or file size is consistent with past unauthorized transactions. If both questions are answered in the affirmative, then a file match may be established

according to block 422. Querying a database for suspect source IP addresses and file types and/or sizes implicated in past unauthorized transactions is generally less processing-intensive than generating and comparing content-based fingerprints.

If examination of the source IP address and file type and/or size do not yield a likely match with a registered copyrighted work, then further assessment criteria using any present metadata or watermarks are preferably applied, according to steps 410-416. If metadata is present in the file according to step 410, and the metadata identifies a registered copyrighted work according to step 412, then a file match is preferably established according to block 422. If either of these questions is answered in the negative, then preferably the following inquiry is whether the file contains a watermark according to step 414. If a watermark is present, and the watermark identifies a registered copyrighted work according to step 416, then a file match may be established according to block 422. Identification by way of metadata or a watermark may be performed by reference to archived data, such as may be stored in the database 244 illustrated in FIG. 4. Inquiring into the presence of metadata or watermark information and querying archived data to compare these identifiers is preferably performed in advance of fingerprinting to achieve desirable speed characteristics if processing resources are limited.

If none of the foregoing assessment criteria indicate the likely presence of a registered copyrighted work, then a content-based fingerprint for a digital sample may be generated according to block 418. But even if one or more of the foregoing assessment criteria indicates a match with a registered copyrighted work, it may be desirable to check at least a portion of the matched results with a fingerprint identification method for validation purposes. That is, each of the foregoing assessment criteria provides only a probability that the unknown content contains a registered copyrighted work. Using fingerprinting techniques to check at least a portion of results matched according to other assessment methods may preferably provide feedback as to the effectiveness of a particular hierarchical identity assessment method.

As noted previously, identification by way of content-based fingerprints is highly accurate, but a primary downside in using fingerprinting is its high consumption of valuable processing resources. Following fingerprint generation, the fingerprint may be compared to an archive of identifiers for registered copyrighted works according to step 420. The archived identifiers may be stored in a database, such as the database 244 illustrated in FIG. 4. If fingerprint comparison identifies a registered copyrighted work according to step 420, then a file match may be established according to block 422. Alternatively, if fingerprint comparison identifies no match according to block 424, then it may be concluded that the digital sample does not correspond to a registered copyrighted work. In such an instance, it is desirable to store the fingerprint in an archive, such as the database 345 illustrated in FIG. 5, to enable retroactive reporting. That is, it may be desirable to monitor transactions in a particular digital work in case an owner of that work later desires to register it with the CPS provider and would like to obtain information regarding transactions in that work pre-dating registration of the work. Depending on the number, frequency, and/or timing of transactions in a particular work, a copyright owner may recognize the benefit of registering the work and/or choose one or more particular business rules to provide an appropriate and desirable level of copyright protection.

When a copyright owner should decide to register a particular work with the CPS provider, one task for the CPS provider is to gather and/or enter potentially useful data corresponding to that work into a database or archive, such as the archive 338 illustrated in FIG. 5. This task may be generally described as data enrichment. Preferably, data enrichment is automated to the extent possible, but manual intervention may be desirable, such as to augment information available to an automated data enrichment service and/or to check and control the quality of automatically entered data. Numerous data fields may be useful in operating a CPS or providing copyright protection services in accordance with the present invention, such as, for example, file name, file size, a content-based fingerprint, commerce artist name, label name, album name, producer name, release date, and others.

FIG. 7 provides an example of a procedure for data enrichment. The first step 500 is to obtain the copyrighted work to be registered in digital form. The CPS provider may obtain digital files, for example, by way of transmission over a network such as the Internet, or by way of a portable digital storage medium such as a CD or DVD. If necessary, the CPS provider may receive an analog copy or a hard copy of a copyrighted work, such as a cassette tape or a photograph, and convert it to digital form. The next step 502 to generate a fingerprint, preferably for each discrete digital work. If an entire music album were provided to the CPS provider, then a separate fingerprint would preferably be generated for each song on that album to facilitate identification of individual songs by the CPS.

A CPS may use Metadata. Inquiry into the presence of owner-supplied metadata may be performed according to step 504. Owner-supplied metadata, which may be found, for example, in a format such as an .id3 or .md5 file associated with the digital work, may be extracted according to block 506. Types of metadata that might be extracted include, for example, artist name, title of the song / movie / work, album name, company / owner name, producer name, release date, and similar information. If no owner-supplied metadata is present, then online metadata archives is preferably queried for the specified copyrighted work according to step 508. Examples of online metadata archives that may be queried for such information include "FreeDB" and "CDDB." If the online archives include metadata for the specified copyrighted work according to block 510, then the metadata is preferably extracted according to step 506 for use in the CPS. If no metadata is available for the work in such a database, then desired information may be added manually according to step 512. Following addition of metadata, any art associated with the work may be added to a CPS database, such as the archive 338 illustrated in FIG. 5. Such associated art may include, for example, an album cover for an audio work, a thumbnail of an image work, or movie art.

Following addition of metadata information and associated art, preferably a query is performed to determine which commercial site or sites, if any, offer the particular copyrighted work for sale according to step 516. Preferably the commercial site(s) are

online websites, and more preferably websites affiliated with the CPS provider such as by contractual affiliation. Address information, preferably in the form of a URL, for commercial websites having the work for sale is then associated with the copyrighted work in a CPS database. A final step may be the addition of a "deep" link (such as a URL) or product code for purchasing the specified registered copyrighted work from the commercial site according to step 518. The foregoing information may be useful in facilitating commercial transactions in registered copyrighted works.

FIG. 8 illustrates an implementation of a CPS 600 utilizing several network appliances 602, 604, 606, 608 distributed along network segments for several watched networks 612, 614, 616, 618. Each watched network 612, 614, 616, 618 connects to a distributed electronic network such as the Internet 620, and each network appliance 602, 604, 606, 608 has access to digital data transmitted between each watched network 612, 614, 616, 618, and the Internet 620. While a network appliance utilized with a CPS generally may operate either in-stream or mirrored along a network segment, the configuration illustrated in FIG. 8 illustrates network appliances 602, 604, 606, 608 configured to receive mirrored data transmitted between watched networks 612, 614, 616, 618 and the Internet 620. Each network appliance is capable of communicating with a CPS network data center 630, which preferably includes such devices as a transaction request broker service 632, a transaction recording and management service 634, a transaction database 636, a raw media storage service 644, and a raw media storage archive 646. The transaction request broker 632 preferably routes and/or manages transactions between various components of the CPS, including various network appliances 602, 604, 606, 608. The transaction database 636 stores information relating to transactions in digital works, with particular emphasis on unauthorized transactions in registered copyrighted works. The transaction recording and management service 634 provides an interface with the transaction database 636. The raw media storage archive 646 may be used to store information including digital works, such as those supplied by copyright owners or duplicated from traffic communicated between a

watched network 612, 614, 616, 618 and the Internet 620. The raw media storage archive 646 may further store fingerprints generated from copyrighted works. The raw media storage service 644 provides an interface with the raw media storage archive 646.

Each network appliance 602, 614, 606, 608 preferably includes a memory for receiving and storing content-based identifiers, including fingerprints. Preferably, each network appliance 602, 614, 606, 608 includes a processor to provide content type identification and content assessment capabilities. Each network appliance 602, 614, 606, 608 may be periodically updated with new identifiers from the network data center 630, such as identifiers for newly registered copyrighted works. The distributed architecture of a CPS according to FIG. 8 facilitates rapid monitoring of high-bandwidth watched networks 612, 614, 616, 618. Each network appliance 602, 604, 606, 608 may communicate with the network data center 630 by way of a public network such as the Internet, a virtual private network, a dedicated private network, or any combination of such connection types to promote system reliability in case one becomes inoperable. Additionally, while FIG. 8 illustrates only a single network appliance at each watched network 612, 614, 616, 618, redundant network appliances may be provided at each location to enhance overall system reliability.

Propagation and utilization of a CPS 100, 200, 300, 600 as disclosed herein enables novel methods of conducting a profitable business enterprise. FIG. 9 illustrates a business method 700 including steps that may be employed according to one or more CPS embodiments. The business method 700 illustrated in FIG. 9 is intended to present merely one example of novel business steps; non-obvious variants omitting certain steps, addition of further steps, and applying disclosed steps in a modified sequence are still contemplated to remain within the scope of the invention.

The first step 702 provided in FIG. 9 is providing and/or maintaining a database (or "registry") of information identifying registered copyrighted works. Herein after a digital work which has been added to the database will be referred to as a "registered work" or

"registered copyrighted work". As new original works are being continuously created and owners of existing copyrighted works or operator of the CPS may elect to protect works by way of a CPS as disclosed herein, a database of identifiers should be designed to grow over time. A data enrichment method, such as that as discussed in conjunction with FIG. 7, is preferably applied to build and maintain the database according to this step 702. A revenue-generating step 704 includes the solicitation of payment in exchange for registering copyrighted works and/or providing copyright infringement protection services. This payment may be solicited by the provider from, for example, copyright owners individually, associations of copyright owners, network providers or operators, or any combination thereof.

Providing copyright protection services according to the present invention generally includes monitoring a network or network segment for transmission of digital signals, as in step 706. Identification of files embodying file types of interest transmitted over the monitored network or network segment may be performed according to step 708. If a file type of interest is found, then one or many of various features may be extracted or generated from the file to facilitate content identification according to step 710. A comparison step 712 is advantageously performed to determine whether the extracted or generated features match one or more identifiers contained in the database maintained according to step 702. If a match is made, then such a match indicates that the file from which the features were obtained contains a registered copyrighted work, as noted in step 714.

A typical follow-up step 716 is to check whether transmission or receipt of the registered copyrighted work has been authorized by the copyright owner. Preferably, the CPS provider maintains a database that identifies authorized senders and/or receivers of digital copyrighted works, and that further includes preferences of the copyright owner for handling transactions in a copyrighted work. Determining whether a particular address is authorized to transmit and/or receive a registered copyrighted work may be performed by querying the database for such information. Regarding handling preferences, such

preferences may be used by the CPS provider to apply business rules to transactions or attempted transactions in registered copyrighted works. For example, some copyright owners such as software developers may distribute copyrighted material according to license agreements that expressly forbid digital transmission of the source code. Such owners might prefer to block all attempted digital transmission of these materials, and communicate this preference to the CPS provider.

If upon application of step 716 it is determined that the transmission is not authorized, then information identifying the transaction may be recorded (such as in the transaction database illustrated in FIG. 8) according to step 718. Recorded information for an unauthorized transaction may include identifiers such as included in FIG. 2. Preferably, transaction reports, such as the report illustrated in FIG. 2, may be generated from some or all of the recorded information. As information contained in transaction reports may be valuable to copyright owners and others, for purposes including but not limited to marketing and seeking licensing revenue, such reports may be sold by the CPS provider in a further revenue generating step 724.

The ability of generating transaction reports and/or blocking content provides additional revenue generation potential by affording businesses and organizations the opportunity to install the CPS on their networks or computers. A per-seat license may be offered to an organization or business to limit and/or monitor the transmission of content by its members and thereby limit the organization's or business' exposure to liability for unauthorized use of content. Similar to the way virus protection software may be installed on individual computers in a local area network of an organization, CPS client software may be installed to afford an organization or business copyright infringement protection.

If transmission of the registered copyrighted work is authorized, then preferably lesser information regarding the transaction may be recorded, such as by incrementing a counter of transactions in the particular registered work, according to step 720. Preferably

less information is recorded in the case of an authorized, lawful transaction to respect the privacy rights of the sender and receiver.

Following recordation of transaction information for an unauthorized transaction according to step 718, business rules may be applied to the transaction according to step 726. As mentioned above, the CPS provider preferably solicits preferences of copyright owners for handling unauthorized transactions in registered copyrighted works, and the CPS provider maintains a database recording those preferences. The preferences are preferably established at the time a work is registered with the CPS, so that business rules to be applied to a particular copyrighted work may be defined before detection by the CPS provider of an unauthorized transaction in a registered copyrighted work. As noted previously, business rules that might be applied include but are not limited to blocking unauthorized transmissions, sending a message to the source address and/or recipient address, referring the source address and/or recipient address to a commercial website, and/or recording transactions in copyrighted works carried by the monitored signal. A further revenue-generating step 728 may follow from the application of business rules, as the CPS provider may solicit payment for referrals to commercial sites, such as websites, where copyrighted works are available for sale and/or commissions for sales resulting from such referrals. Preferably, the CPS provider obtains an affiliation, such as by contract, with commercial sites to provide for referral and/or commission payments. Accordingly, the exemplary business method 700 provided in FIG. 9 provides multiple potential revenue streams to the CPS provider.

Returning to the comparison step 712 wherein the features obtained from a sampled work were compared to identifiers contained in a CPS database, if no match is found, then it may be concluded that the digital sample does not correspond to a registered copyrighted work, as provided in step 730. Yet it may still be useful to record information relating to this work, to facilitate retroactive reporting in case a copyright owner later registers the work with the CPS provider and seeks information relating to its digital distribution. A fingerprint

may be generated from the unregistered work according to step 732. Thereafter, the fingerprint may be stored by the CPS provider in a database or archive such as the database 646 provided in FIG. 8. Preferably, the database (such as database 646 of FIG. 8) is queried to compare the newly generated fingerprint to archived fingerprints for other unregistered works according to step 734. If a match is found from this query, then a transaction counter may be incremented to reflect the number of transactions in the particular work according to step 736. If no match is found, then the fingerprint is preferably added to the database of unregistered works. Regarding the capability of providing retroactive transaction reports, such information may be useful to the copyright owner in selecting particular preferences or business rules to be applied by the CPS provider to protect a copyrighted work following its registration.

As noted previously, U.S. Patent No. 5,918,223 provides a method for performing analysis and comparison of audio data files based upon the content of the data files. An alternative method to that disclosed in U.S. Patent No. 5,918,223 for generating statistical models of digital audio recordings, which are used for probabilistic identification of unknown digital audio streams, is referred to herein as a Stochastic Audio Matching Mechanism (SAMM). If utilized, SAMM is preferably embodied in a software routine that may operated on a device such as a network appliance (e.g., network appliance 104 in FIG. 1, network appliance 204 in FIG. 4, or network appliances 602-608 illustrated in FIG. 8). Discussed below are the mathematical and statistical concepts behind the SAMM system, as well as a description of one implementation of these concepts.

SAMM is a process for generating statistical models of digital audio recordings and using these models for probabilistic identification of unknown digital audio streams. The creation of the models and the identification of unknown audio streams are separate functional processes, but they are logically tied together within the overall goal of audio identification. In practice, the use of SAMM involves the generation of many models for each audio item that is to be identified, and the storage of these models in a SAMM

database. Once the database has been constructed, unknown digital audio streams can be positively or negatively (no match found) identified within a known degree of accuracy using the SAMM database of audio models. SAMM encompasses two distinct functional processes of model generation and audio identification.

It is important to reiterate that the SAMM process is a statistical tool, and that the identification of unknown audio streams is based on the probability that features the unknown audio exhibits matches the features from a known audio sample. A probability over a given threshold likely indicates that the unknown audio stream matches (corresponds) to the current known sample being compared against, while a probability under the given threshold indicates that the unknown audio stream does not match the current model being compared against. Since the system is probabilistic against a threshold, there are no absolutes when using this process.

The model generation process and the audio identification process share a common set of data that they operate upon. These data are the features of the digital audio stream. A single feature is a collection of the representative characteristics of the audio stream at a single point in time (currently, about twenty characteristics per feature). Many features are extracted for a given audio stream, usually one hundred per second. Once the collection of the feature set is completed, SAMM can then generate a model for storage, or use the feature set to compare against known models. The details pertaining to feature extraction, model creation and audio matching are explained fully in the Process Detail section.

A. SAMM Overview

FIG. 10 provides a generalized description of the data flow within SAMM. Boxes 792, 794, 796 represent the major processes of Feature Extraction, Model Generation, and Audio Matching, as will be described herein in further detail. The plain text in FIG. 10 represents the input and outputs for each process. Raw audio input 791 is provided to the feature extraction process 792, which outputs extracted features 793 to both the model generation and audio matching processes 794, 796. The model generation process results in

creation of a model 795, while the audio matching process results in either positive or negative identification 797.

B. SAMM Inputs and Outputs

1. Feature Extraction

a. Data Input

Regardless of the desired SAMM functionality (i.e., model generation or audio stream identification), at least one feature, and preferably a collection of features, is generated from an initial digital audio stream, such as the raw audio data 791 illustrated in FIG. 10. This audio stream is therefore the initial data input for the system as a whole. Raw digital audio 791 coming into the system is preferably first decoded and down-sampled to a pulse code modulation (PCM) stream, such as at a frequency of 16 kHz. Typically, .mp3 and CD quality audio streams are encoded at 44.1 kHz. This decompression and conversion may be performed by a readily available third party utility such as the Linux utility mpg123. Once decompressed and converted, the PCM stream is assembled into a data array, which is the primary input into the Feature Extraction process 792.

b. Parametric Input

The statistical parameters used in feature extraction should be determined before the extraction process 792 occurs. The primary parameters used in the mathematical and statistical formulas used for feature extraction (discussed in further detail, infra) are summarized below with corresponding sample values for illustration purposes:

- Sampling rate of the incoming PCM data (e.g., 16kHz).
- Window length (which is a function of the sample rate).
- Skip rate (which is a function of the sample rate).
- Pre-emphasize constant (e.g., 0.97).
- Filter bank count (e.g., 20) - this is the number of datum in a feature.
- Filter bank channels (e.g., Filter bank count – 1) – number of computed Mel-Frequency Cepstral Coefficient (MFCC).

- Center frequency (e.g., Filter bank count + 2).

These parameters are preferably set or calculated software.

c. Feature Output

The output of the Feature Extraction process 792 is a collection of feature vectors, the number of which is determined by the parametric input to the process. Each vector preferably consists of Filter bank count number of floats and this vector statistically represents the digital audio stream at a particular moment in time. The collection of features is treated within the software as an array of arrays (two-dimensional array) of floats, and this serves as the data input to the 2. Model Generation process 794 and 3. Model Matching process 796.

2. Model Generation

a. Data Input

The input to the Model Generation process 794 is an array of an array of floats (collection of feature vectors 793) representing the audio to be modeled. This is the output of the 1. Feature Extraction process 792.

b. Parametric Input

The statistical parameters used in the extraction of features should be determined before execution of the Feature Extraction process 792. The primary parameters chosen for the mathematical and statistic formulas used in model generation are summarized below with corresponding sample values for illustration purposes:

- Vector length (e.g., Filter bank count).
- Mixture count (e.g., 8).
- Max iterations (e.g., 15).
- Max frames (e.g., 3000 – this corresponds to 100 frames per second for 30 seconds of audio).
- Variance threshold (e.g., 0.001).

These parameters are preferably set or calculated within software.

c. Model Output

A generated model 795 is preferably a binary file containing statistical information about the raw audio 791 from which the original feature set was generated. The output model 795 is preferably stored in a "SAMM" database (such as, for example, the database 338 illustrated in FIG. 5 or the database 646 illustrated in FIG. 8) for use in a model matching process 796.

3. Model Matching

a. Data Input

The input to the model matching process 796 is preferably an array of an array of floats (collection of feature vectors 793) representing the audio to be identified. This is the output of the 1. Feature Extraction process 792.

b. Model Matching Result

Output from the model matching process 796 is preferably a textual representation of the identification result. If the feature set from a digital audio stream did not have a match against any model in a SAMM database, a "NO_MATCH" string may be returned. If the statistical attributes of the digital audio stream compare favorably against a model in a SAMM database, then the string "MATCH - <ID>" may be returned, where "<ID>" may be replaced by a unique SAMM database identification number of the model that the input matched with a degree of certainty.

C. Process Detail

1. Feature Extraction

a. Concept Overview

The primary goal of the feature extraction process 792 is to develop a representation of the acoustic signal suitable for classification. A good set of features for this problem should take into account the properties of the human ear while maintaining a high rate of data compression. Because the ear is most sensitive to changes in spectral magnitude and least sensitive to signal phase difference, the features used here preferably capture the

spectral shape of the signal over small "snap-shots" in time. In particular, the audio may be analyzed over small blocks of samples during which the signal is assumed to be short-time stationary (20 to 25 ms is reasonable for speech and most audio). Overlapping windowed segments of audio may be extracted at a rate of, for example, 100 snap-shots per second to produce a vectored feature stream for classification. Different extraction rates may be used. Each frame of audio consisting of approximately 25 ms of PCM samples (e.g., 400 samples @ 16kHz) may be converted into a multi-dimensional, preferably 20-dimensional, vector that encodes the spectral shape and relative-energy of the signal. The feature vector used in the audio classifier is described in further detail below.

b. Mathematics/Statistics

Observation vectors are computed periodically, preferably every 10 ms, from short-time windowed segments of audio data, preferably 25 ms in length. For each frame of audio, a multi-dimensional, preferably 20-dimensional, observation vector is extracted consisting of multiple Mel Frequency Cepstral Coefficients (MFCC), preferably 19 in number, and one normalized log-energy term. A block diagram of a MFCC feature extraction process 800 is illustrated in FIG. 11. In a first step 802, raw digital audio is decoded and down-sampled to a PCM stream, such as at a 16 kHz frequency. In a second step 804, short-time windowed segments are extracted from the down-sampled stream. According to a third step 806, a sampled frame is windowed. The feature extraction begins by pre-emphasizing the audio to remove glottal and lip radiation effects according to a fourth step 808. The pre-emphasis operation is implemented as a first order Finite Impulse Response (FIR) filter given by

(Eqn. 1) $$H(z)=1-0.97z^{-1}$$

where z represents a one sample delay. Note that in the time-domain, the resulting signal is given by $y(n) = s(n) - 0.97s(n-1)$ where $y(n)$ represents the pre-emphasized signal and $s(n)$ represents the input signal. Next, the magnitude spectrum of the waveform is computed using the Discrete Fourier Transform (DFT) according to step 810. The linear frequency

axis is then warped onto the Mel scale according to step 812 in order to take into account the relationship between frequency and "perceived" pitch. The mapping between the linear frequency scale and Mel scale is given by

(Eqn. 1) $$f_{mel} = 2595\log_{10}\left(1+\frac{f_{linear}}{700}\right)$$

The warped magnitude spectrum is next passed through a bank of triangular-shaped filters that uniformly partition the Mel frequency scale into P regions according to step 814. Note that uniformity on the Mel frequency scale takes into account nonlinear sensitivity of the ear across frequency. For 16 kHz sampled audio, 20 filters (P=20) are used. The filter outputs generate a discrete set of P log-energy terms, $(e[j], j=1..P)$. Let $w_j[k]$ represent the weight of the jth filter to the kth discrete frequency of the sampled signal $s(n)$ and let $|S_{mel}[k]|$ represent the DFT magnitude spectrum of $s(n)$ warped onto the Mel frequency scale. Assuming an N point DFT of the signal, the log-energy within the jth filter bank is given by,

(Eqn. 2) $$e[j]=\log_2\left(\sum_{k=0}^{N-1} w_j[k]\cdot|S_{mel}[k]|\right) \quad \text{for j = 1, 2, ... } P$$

Thereafter, the 19 MFCCs $(c_t[i], i=1..19)$ are computed for each excised frame of audio by decorrelating the filter outputs according to step 816 using the discrete cosine transform (DCT),

(Eqn. 3) $$\tilde{c}_t[i]=\sqrt{\frac{2}{P}}\sum_{j=1}^{P}\left(e[j]\cdot\cos\left(\frac{\pi i}{P}(j-0.5)\right)\right)$$

Finally removing the long-term mean from the features normalizes the MFCC parameters. This process, known as *Cepstral Mean Normalization,* helps to reduce the influence of channel mismatch on the excised features (e.g., in song classification such mismatch can occur when different codecs are used to encode the signal or if frequency equalization is applied during the encoding process). The final 19 MFCCs are given by

(Eqn. 4) $$c_t[i] = \frac{1}{T}\sum_{t=1}^{T}\tilde{c}_t[i]$$

The 19 dimensional vector is augmented with a normalized log-energy component, which is calculated for each frame of data. Finally, the log-energy term is calculated by first taking the log of the sum of the squared data samples. Let $s_t(n)$ represent the nth sample from the tth excised frame of audio. Assuming Ns samples per frame of audio, an initial frame-based energy term is computed as follows,

(Eqn. 5) $$\tilde{e}_t = \log_2\left(\sum_{n=1}^{Ns} s_t^2(n)\right)$$

The energy outputs are normalized to range between –5.0 and +1.0 and are augmented as the 20th feature vector element.

c. Feature Extraction Implementation

The implementation discussions herein are intended is to provide a high-level mapping of the concepts and the mathematics for providing SAMM functionality sufficient to enable one skilled in the art to practice the inventive method. In furtherance of this goal, FIGS. 12a - 12b provide annotated sequence diagrams to additionally detail the program and data flow of the individual processes. The following sections are intended to discuss the inner workings of SAMM in the context of the concepts and mathematics at the object level.

The implementation of SAMM is preferably performed in an object-oriented fashion, such as in the C++ programming language, thus all objects described in this section and in the following sections are C++ objects.

A higher level calling entity 902 within the Media Analysis System initiates the process of feature extraction. A utility object within SAMM called the AudioAnalyzer 904, which is instantiated and managed by the calling entity, performs the actual extraction. The calling entity 902 is also responsible for managing the collection of raw data from which features are extracted, and managing the AudioAnalyzer's use of this data. The calling

entity executes various member functions on the AudioAnalyzer 904 with the ultimate goal being the extraction of features from raw audio data and the storage of this collection of features as a data member within the AudioAnalyzer object 904. Once populated with the features, the AudioAnalyzer object 904 is used as the storage and wrapper of the features as they are used in the process of model generation of audio identification.

2. Media Model Generation

a. Concept Overview

It is assumed that perceptual difference between audio music and other audio media are primarily manifested by the characteristics of the signal's spectrum. This is illustrated in FIGS. 13a-13b for two segments of audio from music pieces. Here, frequency is plotted along the y-axis while time is plotted along the x-axis. In FIG. 13a we see distinct sequences of spectral patterns emerge as the song progresses from piano key strokes through a percussion hit through finally a sequence of synthesizer key strokes. In FIG. 13b we see other patterns manifested when a singer vocalizes a word in the song. In FIG. 13a a piano keystroke leads into a percussion followed by synthesizer notes. In FIG. 13b an artist sustains vocalization while singing. It is assumed that the sequences of spectral patterns across the time-sequence of the audio represent the signature or "footprint" of the song. Modeling the spectral characteristics of each audio clip allows one to distinguish between artists and music pieces.

Ideally, one would prefer to model the trajectory of the spectral events in order to capture the evolution of the audio over time. However, it is pointed out that the explicit modeling of temporal events leads to a classification algorithm which is susceptible to performance degradations when the signal is shifted in time.

The proposed algorithm considered here assumes that the sequence of features extracted from the song is statistically independent. Under this assumption, the likelihood of observing a feature vector $\vec{x}_t$ at time t is not dependent on the feature vector $\vec{x}_{t-1}$ extracted at time $t-1$ or any other time for that matter. In other words, the likelihood of observing

sequence of T feature vectors, $X = \{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_T\}$ given a model λ for an audio segment can be expressed as,

(Eqn. 6) $$p(X \mid \lambda) = \prod_{t=1}^{T} p(\vec{x}_t \mid \lambda)$$

Eqn. 6 states that the likelihood of observing a sequence of feature vectors given a particular model for a music clip is based on the product of the individual likelihood of observing each feature vector excised from the clip. At 100 feature vectors per second of audio, complete computation of Eqn. 6 for 30 seconds of a song would require the product of *T=3000* individual likelihoods. Note that since repeated multiplication of many numbers smaller than 1 leads to numerical underflow problems on most PC hardware. Therefore, the likelihood in Eqn. 6 is generally expressed in terms of its log-likelihood,

(Eqn. 7) $$\log p(X \mid \lambda) = \sum_{t=1}^{T} \log p(\vec{x}_t \mid \lambda)$$

The basic concept behind the audio modeling scheme is that each song under consideration can be modeled by characterizing the *statistical distribution* of the feature vectors excised from an example of the song. In doing so, the audio modeling scheme becomes less sensitive to slight alterations in the features. Such alterations can be experienced due to differences in audio codecs, time-shifts in the signal, sampling rate, etc. Unlike audio "fingerprinting" schemes that try to find an *exact match* of the audio to a known model, the statistical approach returns the *likelihood* or probability that the observed set of features were generated by a model, λ. Therefore given a set of *S* modeled songs, $\{\lambda_1, \lambda_2, \ldots \lambda_S\}$, and an unknown audio clip with excised feature sequence, $X = \{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_T\}$, the goal of the search is to find the model λ_s with the maximum likelihood of generating the unknown feature sequence. The song associated with this model is assumed to best match the unknown. In other words,

(Eqn. 8) $$\lambda_s = \arg\max_{1 \le s \le S} \{\log p(X \mid \lambda_s)\}$$

Of course, Eqn. 8 assumes that the feature sequence $X = \{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_T\}$ was generated from at least one of the known S songs in the database. A case in which the test sequence is outside of a known database will be considered, *infra*.

b. Mathematics/Statistics

It is assumed that the feature vector sequence $X = \{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_T\}$ is statistically independent and generated from a random process consisting of a linear combination of Gaussian basis functions. Models of this type are known as Gaussian Mixture Models (GMMs). GMMs have been used in the past for problems such as Speaker Identification and Language Identification. A Gaussian Mixture Model characterizes the likelihood of observing a feature vector $\vec{x}$ as a weighted combination of Gaussians:

i. The Gaussian Mixture Model

(Eqn. 9) $$p(\vec{x} \mid \lambda) = \sum_{m=1}^{M} w_m \cdot b_m(\vec{x})$$

where $b_m(\vec{x})$ is the multivariate Gaussian density. For a D-dimensional feature vector, $b_m(\vec{x})$ can be expressed as,

(Eqn. 10) $$b_m(\vec{x}) = \frac{1}{(2\pi)^{D/2} |\Sigma_m|^{1/2}} \exp\left\{-\frac{1}{2}(\vec{x} - \vec{\mu}_m)' \Sigma_m^{-1} (\vec{x} - \vec{\mu}_m)\right\}$$

Here, $\vec{\mu}_m$ and Σ_m represents the vector mean and covariance of the mth Gaussian density respectively. Further, the weights for the Gaussian functions follow the sum-to-one property,

(Eqn. 11) $$\sum_{m=1}^{M} w_m = 1$$

For data sparsity and speed issues, the covariance matrix in the model is assumed to be diagonal, i.e., all elements off the diagonal are zero-valued. Therefore, our model consists of M mixture weights, mean vectors, and covariance matrices. Typically numbers of mixtures needed to accurately model a song range between M=10 and M=32.

ii. Parameter Estimation

Estimation of the model parameters is based on the Expectation-Maximization (EM) algorithm A. Dempster, N. Laird, and D. Rubin, "Maximum Likelihood from Incomplete Data Via the EM Algorithm," *J. Royal Stat. Soc.*, Vol. 39, pp. 1-38, 1977, and L. Baum et al., "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains," Ann. Math. Stat., Vol. 41, pp. 164-171, 1970, both references of which are incorporated by reference as though fully set forth herein. A practical application of the update equations can be found in D. Reynolds, R. Rose, "Robust Text Independent Speaker Identification Using Gaussian Mixture Speaker Models," *IEEE Transactions on Speech and Audio Processing*, Vol. 3, No. 1, pp. 72-83,Jan. 1995, which is incorporated by reference as though fully set forth herein. The parameter estimation algorithm is iterative. At each iteration, a new set of model parameters are determined which increase the total likelihood of the training patterns against the current model. In general between 6-10 iterations of the parameter update equations are required before model convergence.

(1) Initialization: The M mean vectors of the model are initialize to randomly chosen data vectors in the training set of T vectors, $X = \{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_T\}$. The M covariance vectors are initialized to have unit variance for each feature element and mixture weights are initialized to have equal weighting (i.e., $w_m = 1/M$).

(2) *Iterative Update:* Assuming diagonal covariance matrices, the observation probability, $b_m(\vec{x}_t)$, can be expressed by,

(Eqn. 123)

$$b_m(\vec{x}_t) = \sum_{m=1}^{M} \frac{w_m}{(2\pi)^{D/2} \sqrt{\prod_{j=1}^{D} \sigma_m^2[j]}} \exp\left\{ -\frac{1}{2} \sum_{j=1}^{D} \frac{(x_t[j] - \mu_m[j])^2}{\sigma_m^2[j]} \right\}$$

(3) *Likelihood:* Let $p(m \mid \vec{x}_t, \lambda)$ represent the *a posteriori* probability of the *m*th modeled Gaussian given feature vector $\vec{x}_t$,

(Eqn. 14) $$p(m \mid \vec{x}_t, \lambda) = \frac{w_m \cdot b_m(\vec{x}_t)}{\sum_{k=1}^{M} w_k b_k(\vec{x}_t)}$$

The update equations for the mixture weights, mean vectors, and diagonal-covariance matrices can then be expressed as,

Mixture weight update

(Eqn. 135) $$w_m = \frac{1}{T} \sum_{t=1}^{T} p(m \mid \vec{x}_t, \lambda)$$

Mean vector update

(Eqn. 146) $$\vec{\mu}_m = \frac{\sum_{t=1}^{T} p(m \mid \vec{x}_t, \lambda) \cdot \vec{x}_t}{\sum_{t=1}^{T} p(m \mid \vec{x}_t, \lambda)}$$

Diagonal-Covariance update

(Eqn. 157) $$\bar{\sigma}_m^2 = \frac{\sum_{t=1}^{T} p(m \mid \vec{x}_t, \lambda) \cdot \vec{x}_t^2}{\sum_{t=1}^{T} p(m \mid \vec{x}_t, \lambda)} - \bar{\vec{\mu}}_m^2$$

(4) *Check Likelihood:* The total likelihood ((Eqn. 7) of the data iteration *i* should be greater than that at iteration *i-1*. Note that over-iterating can reduce the performance of the classifier.

iii. Practical Considerations

There are several practical ranges for the parameters that can be observed during model estimation in order to determine whether or not the convergence criteria for the iterative EM algorithm are satisfied. While absolute criterion that total likelihood of the data against model should increase at each iteration, the following parameter ranges should be maintained,

$$0 \le \bar{w}_m \le 1$$
$$\bar{\sigma}_m^2 > 0$$
$$0 \le b_m(\vec{x}_t) \le 1$$

iv. Notes on Algorithmic Efficiency for Likelihood Calculations

Computation of the likelihood of an individual feature vector against a known model is generally expressed in the log-domain to avoid numerical underflow problems,

(Eqn. 18) $$\log p(\vec{x} \mid \lambda) = \log\left\{\sum_{m=1}^{M} w_m \cdot b_m(\vec{x})\right\}$$

As mentioned, we can expand Eqn. 18 by inserting Eqn. 13 for $b_m(\vec{x}_t)$:

(Eqn. 19) $$\log p(\vec{x} \mid \lambda) = \log\left\{\sum_{m=1}^{M} \frac{w_m}{(2\pi)^{D/2} |\Sigma_m|^{1/2}} \exp\left\{-\frac{1}{2}(\vec{x} - \vec{\mu}_m)' \Sigma_m^{-1} (\vec{x} - \vec{\mu}_m)\right\}\right\}$$

Assuming diagonal covariance matrices, Eqn. 19 becomes,

(Eqn. 20) $$\log p(\vec{x} \mid \lambda) = \log\left\{\sum_{m=1}^{M} \frac{w_m}{(2\pi)^{D/2}\sqrt{\prod_{j=1}^{D}\sigma_m^2[j]}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{D}\frac{(x_t[j]-\mu_m[j])^2}{\sigma_m^2[j]}\right\}\right\}$$

Evaluation of Eqn. 20 requires M exp operations, 3D+M multiplies, and one log operation. In general, we observe that one Gaussian tends to dominate the likelihood computation. Therefore, if it is assumed that only one Gaussian contributes significantly and the remaining M-1 Gaussians have zero-probability, it can be shown that the expression in Eqn. 20 can be approximate as follows,

(Eqn. 21) $$\log p(\vec{x} \mid \lambda) \approx \arg\max_{1 \le m \le M}\left\{C_m - \frac{1}{2}\sum_{j=1}^{D}\frac{(x_t[j]-\mu_m[j])^2}{\sigma_m^2[j]}\right\}$$

Here C_m is a mixture-density dependent constant that can be pre-computed at run-time,

(Eqn. 22) $$C_m = \log(w_m) - \frac{D}{2}\log(2\pi) - \frac{1}{2}\sum_{j=1}^{D}\log(\sigma_m^2[j])$$

Further computational savings for Eqn. 21 can be obtained using *partial distance elimination (PDE)* and *feature component reordering (FCR)* as described in B. Pellom, R. Sarikaya, J. Hansen, "Fast Likelihood Computation Techniques in Nearest-Neighbor based search for Continuous Speech Recognition," *submitted to IEEE Signal Processing Letters.* The basic idea of partial distance elimination is to compute Eqn. 21 for the first mixture Gaussian (m=1) in its entirety and only partially compute Eqn. 21 for the remaining mixtures. Note here that since Eqn. 21 seeks to determine the mixture component which maximizes the expression on the left-hand-side (LHS) of the equation, the summation over the D vector elements can be prematurely stopped as soon as the partial accumulation falls below that of the best-scoring mixture. The end result is that we compute the entire equation for at least one of the Gaussian basis functions but only partially compute the expression for some or all remaining mixtures. The PDE algorithm is guaranteed to give the same output value as the

complete computation of Eqn. 21 (i.e., if Eqn. 21 were to be computed as shown). Alone, PDE reduces the computation by 10% based on empirical simulations.

The effectiveness of the PDE algorithm can be enhanced when combined with *feature component reordering (FCR).* FCR seeks to re-order the sequence of features computed in the summation term in Eqn. 21 such that the partial summation more quickly approximates the true value of likelihood computed over all the elements. The re-ordering of the feature sequence (i.e., $j \Rightarrow f(j)$) is determined empirically from observed data. FCR combined with PDE reduces the computation of Eqn. 21 by 30% based on empirical simulations. Note that PDE and FCR both assume that the "nearest-neighbor" approximation for log-likelihood calculations is used.

c. Model Generation Implementation

FIG. 14 is an annotated sequence diagram describing the process of model generation within SAMM. A calling entity 902 initiates model creation via the use of the AudioModeler object 906. The inputs required for the AudioModeler object 906 are an AudioAnalyzer object 902, which contains the set of features to be modeled, and a reference to the model to be created. This reference is passed to the AudioModeler object 906, and the model is created in-situ.

3. Media Identification

a. Concept Overview

The goal of the media identification algorithm is decide whether or not the audio material under test matches one of the S songs modeled by the system. If the system decides that the audio is from one of the modeled songs in the database, the identifier must provide a classification of which song the material is from.

b. Mathematics/Statistics

The media identification task can be cast as a binary hypothesis problem. Under hypothesis H1 we conclude that the audio under consideration was emitted from one of the known models λ_s (s=*1*..*S*). Under hypothesis H0, we conclude that the audio was not

emitted from any of the known modeled songs. The optimal processor for the binary hypothesis problem is the likelihood-ratio test,

$$\text{(Eqn. 23)} \qquad \frac{p(X \mid \lambda_s)}{p(X \mid \lambda_{\bar{s}})} \underset{H0}{\overset{H1}{\geq}} \Theta$$

In other words, we compare the ratio of probabilities that the feature sequence X was emitted from known model λ_s against the probability that the feature sequence was emitted from an unknown source $\lambda_{\bar{s}}$ (i.e., a song not in the database). The resulting ratio is compared to a decision threshold Θ. If the ratio falls below the threshold, we conclude hypothesis H0, otherwise we conclude hypothesis H1. In the log-domain, the log-likelihood ratio processor becomes,

$$\text{(Eqn. 24)} \qquad \underbrace{\log p(X \mid \lambda_s)}_{\substack{\text{obtained from song} \\ \text{in finite database}}} - \underbrace{\log p(X \mid \lambda_{\bar{s}})}_{\substack{\text{obtained from} \\ \text{song outside of} \\ \text{modeled database}}} \underset{H0}{\overset{H1}{\geq}} \log\Theta$$

It is clear that the first term on the LHS of Eqn. 24 can be expressed as a linear combination of Gaussian basis functions, estimated from the song under consideration for the test. However, the model $\lambda_{\bar{s}}$ that characterizes the H0 hypothesis is not so clearly defined. Currently, our solution is to model $\lambda_{\bar{s}}$ using the top N nearest models to X excluding λ_s. Eqn. 24 becomes,

$$\text{(Eqn. 25)} \qquad \underbrace{\log p(X \mid \lambda_{s=1})}_{\substack{\text{obtained from song} \\ \text{in finite database}}} - \underbrace{\frac{1}{N}\sum_{n=2}^{N+1} \log p(X \mid \lambda_{s=n})}_{\substack{\text{now obtained from} \\ \text{songs inside of} \\ \text{modeled database}}} \underset{H0}{\overset{H1}{\geq}} \log\Theta$$

Here, $\lambda_{s=1}$ is used to denote the model with the highest-likelihood for the unknown test observation sequence X and $\lambda_{s=n}$ (n=2..N+1) is used to denote the next N top scoring models for same test observation sequence. The goal here is that the model for condition H0 should model the case of "any" song present while the first term in Eqn. 25 should model the case of a "particular" song we are interested in. In the next section, we consider the search

mechanism. Note that the second normalization term in Eqn. 25 has been used for problems such as Speaker Voice Verification. This technique is sometimes referred to as "cohort normalization".

i. Search Algorithm

Given an unknown sequence of T excised feature vectors, $X = \{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_T\}$, and known modeled set of S song titles with associated GMMs $\{\lambda_1, \lambda_2, \ldots \lambda_S\}$, the search is defined as follows:

(1) Initialization: Initialize the accumulated log-likelihoods *C[s]* of all song models to zero. All songs are assumed to be active and potential candidates for *X*.

(2) Update: For each active song model, pickup one feature vector from the stream at time instant *t* and update the log-likelihood of each song model,

$$C[s] = C[s] + \log p(\vec{x}_t \mid \lambda_s)$$

(3) Prune: After a sufficient block of features have been examined, prune a fraction of the remaining models that have the lowest log-likelihood score C[s]. If fewer than *N*+1 models remain, do not prune and models (since they are required to compute Eqn. 25).

(4) *Repeat:* Repeat steps 2 and 3 until *N*+1 models remain or all feature vectors in the stream have been consumed.

ii. Verification Algorithm

Utilizing the *N*+1 models with the largest log-likelihood, we hypothesize that the model with the absolute highest likelihood is the song representing the unknown feature sequence. We test this hypothesis using the likelihood ratio test (Eqn. 25). If the computed log-likelihood ratio falls below the threshold, we assume that the unknown is not a modeled

song in our database. Otherwise, the best matching model (i.e., the one with the highest likelihood) is assumed to be the song that represents the unknown (our match).

c. Media Identification Implementation

FIG. 15 is an annotated sequence diagram describing the process of media identification within SAMM. The implementation of the 3. Media Identification process is similar to 1. Feature and 2. Media Model Generation. A calling entity 902 initiates the identification process via the use of the AudioSearch object 908. The inputs required for the AudioSearch object 902 are an AudioAnalyzer object 904, which contains the set of features to be searched, and a reference to the in-memory database used to store all the known models against which SAMM is comparing the unknown audio input.

Accordingly, novel systems and methods for protecting digital works have been disclosed. While embodiments and applications of the invention have been shown and described, it would be apparent to those skilled in the art that many more modifications are possible without departing from the inventive concepts herein. The invention, therefore, is not to be restricted except in the spirit of the appended claims.

What is claimed is:

1. A method of identifying transmissions of digital works comprising:

maintaining a registry of information permitting identification of digital works;

monitoring a network for transmission of at least one packet-based digital signal;

extracting at least one feature from the at least one digital signal;

comparing the extracted at least one feature with registry information to assess whether the at least one digital signal includes a transmission of at least one portion of a registered digital work; and

taking action based on the result of the comparison

2. The method of claim 1, further comprising the step of initiating an action based on the result of the comparison step, wherein the action is in accordance with at least one pre-defined business rule(s).

3. The method of claim 1, wherein said action comprises the step of recording information identifying transmissions of digital works.

4. The method of claim 1, further comprising the step of extracting the recipient address from the digital transmission and determining whether an address is authorized to receive a registered work included in the at least one digital signal.

5. The method of claim 4, further comprising the step of initiating an action based on the result of authorization determination step, wherein the action is in accordance with pre-defined business rules.

6. The method of claim 1, further comprising the step of extracting the source address from the digital transmission and determining whether an address is authorized to transmit a registered work included in the at least one digital signal.

7. The method of claim 6, further comprising the step of initiating an action based on the result of authorization identification step, wherein the action is in accordance with pre-defined business rules.

8. The method of claim 2, wherein the at least one business rules is selected from the set consisting of: interrupting transmission of the at least one digital signal; providing a message to the recipient address; providing a message to the source address; referring the recipient address to a commercial site where the registered digital work may be purchased; forwarding an advertisement to the recipient address; recording information about the transmission; and reporting information about the transmission.

9. The method of claim 2, wherein the at least one business rules is selected from the set consisting of: inserting different digital content into the transaction; or replacing the digital content with other digital content.

10. The method of claim 1, wherein the at least one digital signal comprises one or more of digital audio information, digital video information, digital text, software, or digital image information.

11. The method of claim 1, further comprising the step of preparing reports including information on transmissions of digital works.

12. The method of claim 1, further comprising the step of network traffic analysis.

13. The method of claim 1, wherein the at least one feature extracted from the digital signal is selected from the group consisting of: file name, file size, file type, source address, recipient address, metadata identifiers, a watermark, a file hash, protocol type, text content, and enough data to generate a content-based fingerprint.

14. The method of claim 1, further comprising the step of generating a content-based fingerprint from the at least one digital signal.

15. The method of claim 1, wherein the registry includes content-based fingerprints which represent digital works, and the comparison step includes the comparison of fingerprint(s) generated from the at least one digital signal with fingerprints in the registry.

16. The method of claim 1, wherein the assessment of whether the comparison of at least one extracted feature with registry information is performed according to a hierarchy of comparison criteria.

17. The method of claim 1, wherein the registry is composed of information pertaining to copyrighted works.

18. The method of claim 1, wherein the action taken is to implement a strategy to cache certain digital content in the network .

19. The method of claim 18, wherein the action taken is based on geographic or network topologic information associated with the destination of the transmission.

20. The method of claim 1, wherein the at least one portion comprises information that represents the entire digital work.

21. A digital works identification system comprising:

a first memory for storing information permitting identification of digital works;

a digital input receiver for receiving at least one packet-based digital signal transmitted over a network;

a second memory for storing at least one portion of the at least one packet-based digital signal; and

at least one processor for obtaining at least one feature from the at least one digital signal and comparing the at least one feature with the stored identifying information.

22. The method of claim 21, wherein the system is intended to protect copyrighted works.

23. The method of claim 21, wherein the system is intended to identify new digital works.

24. The system of claim 21, wherein the at least one feature includes data sufficient to generate a content-based fingerprint from the content of the at least one digital signal.

25. The system of claim 21, wherein the at least one feature includes feature vectors generated from the content of the at least one digital signal.

26. The system of claim 21, wherein the at least one digital signal comprises one or more of digital audio information, digital video information, digital representation of text, software, and digital image information.

27. The system of claim 21, wherein the at least one processor identifies digital works carried by the at least one packet-based digital signal.

28. The system of claim 21, wherein the at least one packet-based digital signal is characterized by at least one of: a source address; or a destination address.

29. the system of claim 21, further comprising a third memory for recording the identified transmissions.

30. The system of claim 29, wherein the third memory further includes a transaction database.

31. The system of claim 29, further comprising a report generator for generating reports on transmissions carried by the at least one packet-based digital signal.

32. The system of claim 28, further comprising a transaction request broker for identifying whether a destination address is authorized to receive the content of the at least one digital signal.

33. The system of claim 32, further comprising means for disabling transmission of at least a portion of the at least one packet-based digital signal if the destination address is not authorized to receive the content of the at least one digital signal.

34. The system of claim 28, further comprising a transaction request broker for identifying whether a source address is authorized to transmit the content of the at least one digital signal.

35. The system of claim 34, further comprising means for disabling transmission of at least a portion of the at least one packet-based digital signal if the source address is not authorized to transmit the content of the at least one digital signal.

36. The system of claim 28, further comprising a message generator for sending a message to the source address or destination address upon detection of an identified transmission.

37. The system of claim 36, wherein the message directs a destination address to a commercial site where authorization to obtain the desired content may be purchased.

38. The system of claim 28, wherein other content is inserted into the transmission.

39. The system of claim 28, wherein other content is substituted into the transmission.

40. The system of claim 21, wherein the at least one processor includes:

a content type recognizer for identifying the file type of a packet-based digital signal.

41. The system of claim 21, wherein the at least one processor includes:

a content identifier for identifying the content of a packet-based digital signal.

42. A digital works identification system comprising:

a first database containing features of digital works;

a digital input receiver for receiving at least one digital signal characterized by a source address and a recipient address;

a feature extractor for extracting features from the digital signal; and

a feature comparator for comparing features extracted from the digital signal against features contained in the database.

43. The system of claim 42, wherein the database further contains authorization indicators indicating whether an intended recipient is authorized to receive particular content.

44. The system of claim 42, wherein the database further contains authorization indicators indicating whether an intended source is authorized to transmit a particular digital work.

45. A method for conducting a business enterprise, the method comprising the steps of:

maintaining a first database for storing features permitting the identification of registered digital works;

monitoring a network for transmission of at least one digital signal characterized by a source and recipient address;

extracting at least one feature from the at least one digital signal;

comparing the extracted at least one feature with features from the database to determine whether the digital signal includes a registered digital work; and

taking action based on business rules.

46. The method of claim 45 wherein said taking action comprises recording information identifying transmissions of registered digital works.

47. The method of claim 45, further comprising the step of determining whether a recipient address is authorized to receive transmission of a registered digital work.

48. The method of claim 45, further comprising the step of initiating an action based on the result of the identification step, wherein the action is in accordance with pre-defined business rules.

49. The method of claim 45, further comprising the step of generating at least one feature of a registered digitial work, wherein the at least one feature includes feature vectors.

50. The method of claim 45, wherein the at least one feature includes source address.

51. The method of claim 45, wherein the at least one feature includes destination address.

52. The method of claim 45, wherein the at least one digital signal comprises one or more of audio information, video information, image information, digital representation of text, or software.

53. The method of claim 45, further comprising the step of soliciting owners of digital works to register digital works with the business enterprise for inclusion in the first database.

54. The method of claim 45, further comprising the step of obtaining value from owners of digital works in exchange for registration of digital works in the first database.

55. The method of claim 45, further comprising the step of obtaining value from network providers or network operators in exchange for providing a copyright protection system or copyright protection services.

56. The method of claim 47, further comprising the step of referring an intended recipient not authorized to receive a registered digital work to a commercial site where authorization to receive the registered digital work may be purchased.

57. The method of claim 56, further comprising the step of soliciting value from commercial sites to which intended recipients are referred by the business enterprise.

58. The method of claim 47, further comprising the step of precluding an unauthorized recipient address from receiving uninterrupted transmission of a registered digital work.

59. The method of claim 45, further comprising the step of determining whether a source address is authorized to transmit a registered digital work.

60. The method of claim 59, further comprising the step of precluding an unauthorized source address from sending uninterrupted transmission of a registered digital work.

61. The method of claim 47, further comprising the step of transmitting a message to an unauthorized recipient address.

62. The method of claim 59, further comprising the step of transmitting a message to an unauthorized source address.

63. The method of claim 45, further comprising the step of inserting digital content into the transmission.

64. The method of claim 45, further comprising the step of substituting digital content into the transmission.

65. The method of claim 61, wherein the message includes instructions by which an unauthorized recipient address may purchase authorization to receive a registered digital work.

66. The method of claim 45, wherein the identifiers of registered digital works are generated from the content of the digital works.

67. The method of claim 45, further comprising the step of maintaining a second database recording transmissions of digital works over the network.

68. The method of claim 67 wherein the second database includes recipient addresses.

69. The method of claim 67, wherein the second database includes source addresses.

70. The method of claim 67, further comprising the step of preparing reports including transaction information.

71. The method of claim 67, further comprising the step of selling the reports.

72. The method of claim 45, wherein the network includes a connection to the Internet.

73. The method of claim 45, further comprising the step of maintaining a second database recording information identifying transmissions in digital works over the network.

74. A method for altering the behavior of a device based on the transmissions of digital works comprising:

maintaining a registry of information permitting identification of digital works;

monitoring a network for transmission of at least one packet-based digital signal;

extracting at least one feature from the at least one digital signal;

comparing the extracted at least one feature with registry information to assess whether the at least one digital signal includes a transmission of a registered digital work;

accumulating statistical measurements of the features of the transmitted digital works; and

altering the functional behavior of the device based on the comparison of the transmitted work

75. The method of claim 74, wherein the device is a network component.

76. The method of claim 74, wherein the functional behavior that is altered is allocating bandwidth to different categories of network use.

77. The method of claim 74, wherein a database of frequently identified digital works is maintained such that these works could be inserted into the digital stream in place of the original source transmission.

78. A registry of digital works comprising:

an identifier to link a digital work to associated information about the work;

at least one feature of the digital work permitting the digital work to be identified;

business rules that indicate how the digital work may be transmitted over a digital stream.

79. The method of claim 78, wherein the at least one feature extracted from the digital signal is selected from the group consisting of: file name, file size, file type, source address, recipient address, metadata identifiers, a watermark, a file hash, protocol type, text content, and a content-based fingerprint.

80. The method of claim 78, wherein the business rules indicate whether a particular address or range of addresses is authorized to transmit said digital work.

81. The method of claim 78, wherein the business rules indicate whether a particular address or range of addresses is authorized to receive said digital work.

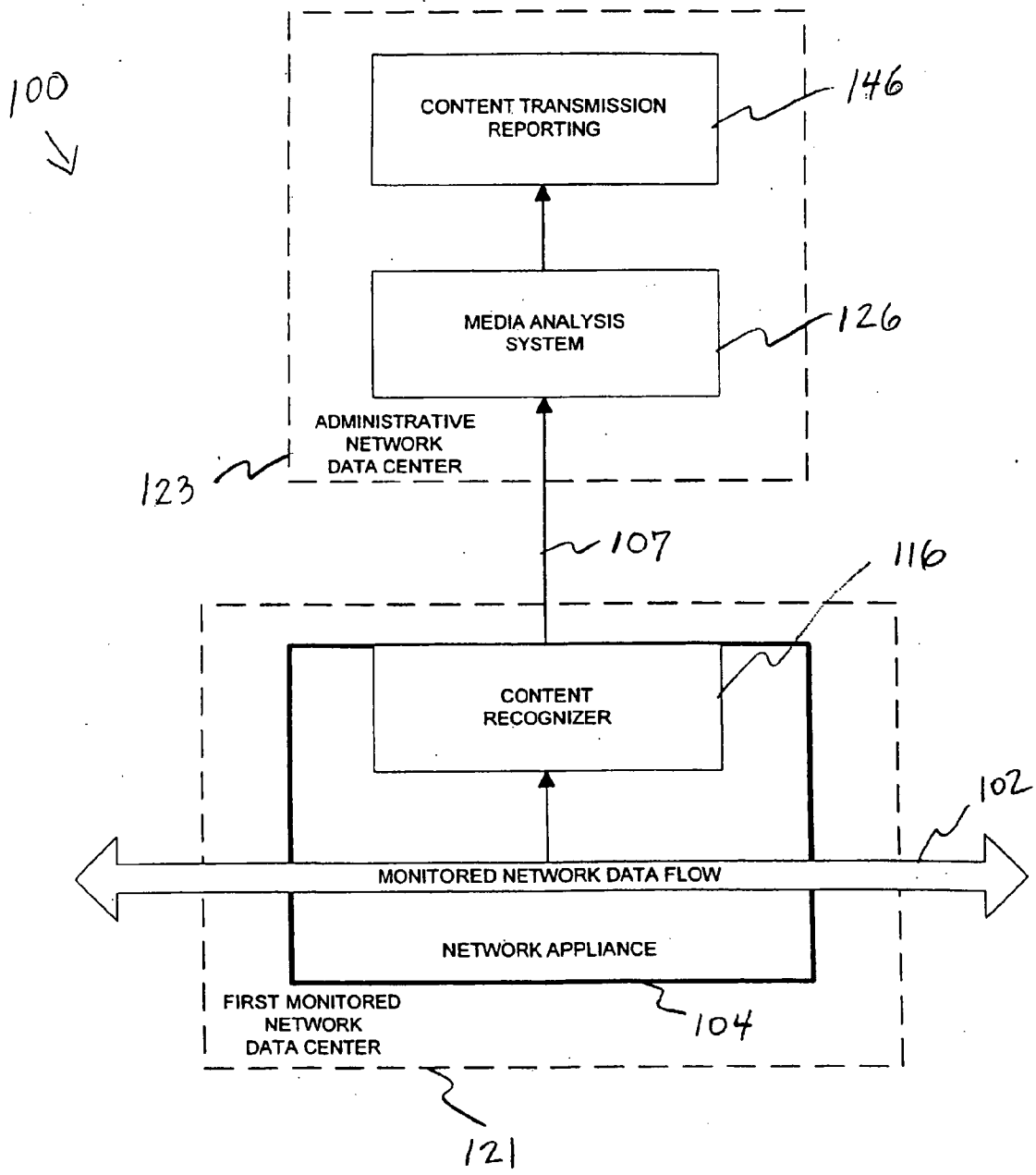

100
CONTENT TRANSMISSION REPORTING
146
MEDIA ANALYSIS SYSTEM
126
ADMINISTRATIVE NETWORK DATA CENTER
123
107
116
CONTENT RECOGNIZER
102
MONITORED NETWORK DATA FLOW
NETWORK APPLIANCE
104
FIRST MONITORED NETWORK DATA CENTER
121

FIG. 1

Source IP Address	Destination IP Address	Date Transmitted	Time Transmitted	File Type
Source Addr. 1	*Source Addr. 2*	*03/01/01*	*09:12:15*	*.mpg*
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
Source Addr. 99	*Source Addr. 99*	*03/06/01*	*11:21:45*	*.avi*

Content / Media Name	Artist Name	Album Name	Record Label / Studio	Meta-Data
Song 1	*Artist 1*	*Album 1*	*Label 1*	*Producer 1*
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
Movie 12	*Artist 6*	*N/A*	*Studio 99*	*Distributor 99*

Unauthorized Count	Blocked Count	Redirected Count	Redirected Purchases	<Other Information>
7	7	2	1	--
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
1	1	1	0	--

FIG. 2

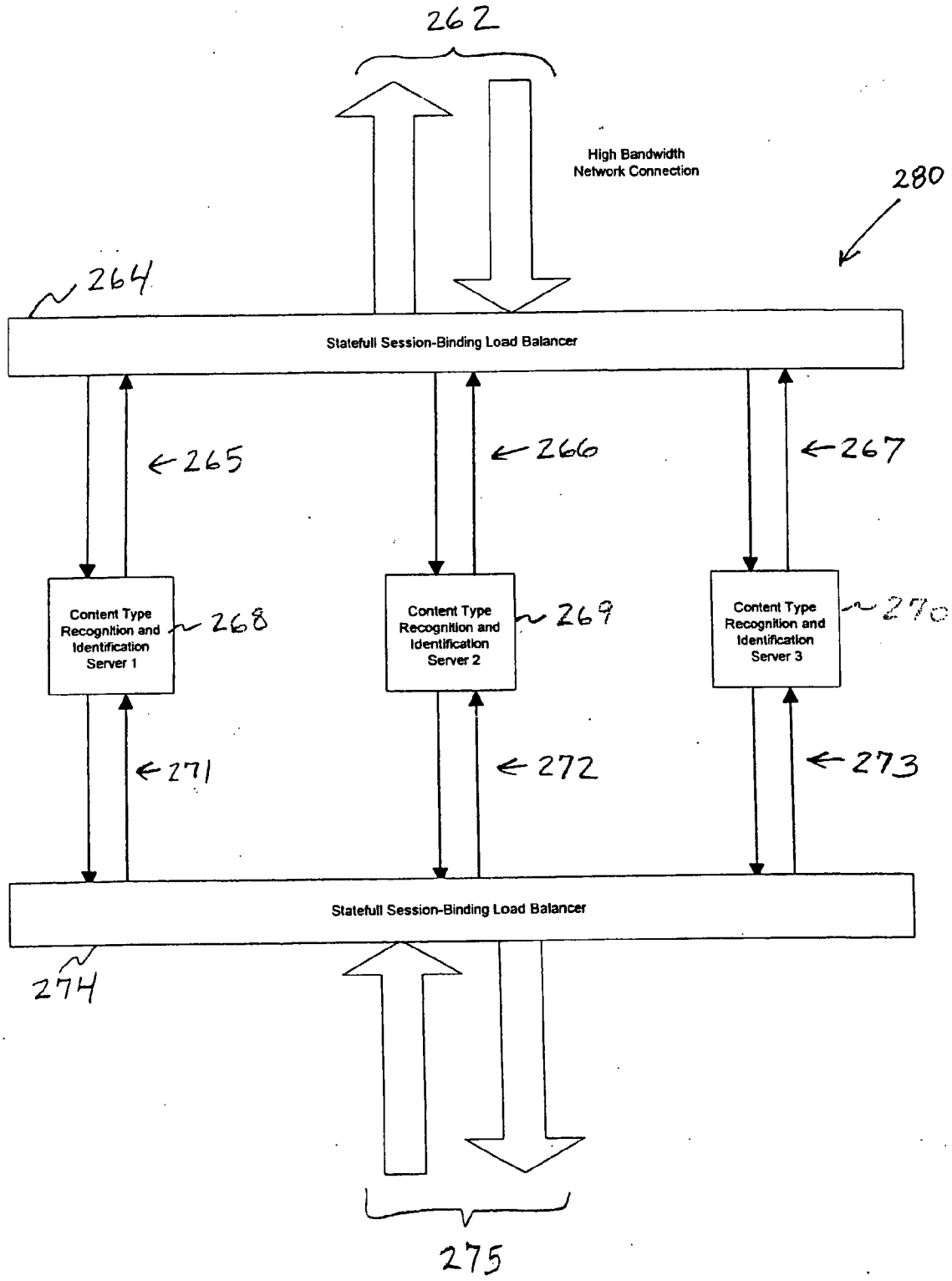

262
High Bandwidth
Network Connection
280
264
Statefull Session-Binding Load Balancer
265
266
267
Content Type
Recognition and
Identification
Server 1
268
Content Type
Recognition and
Identification
Server 2
269
Content Type
Recognition and
Identification
Server 3
270
271
272
273
Statefull Session-Binding Load Balancer
274
275

FIG . 3

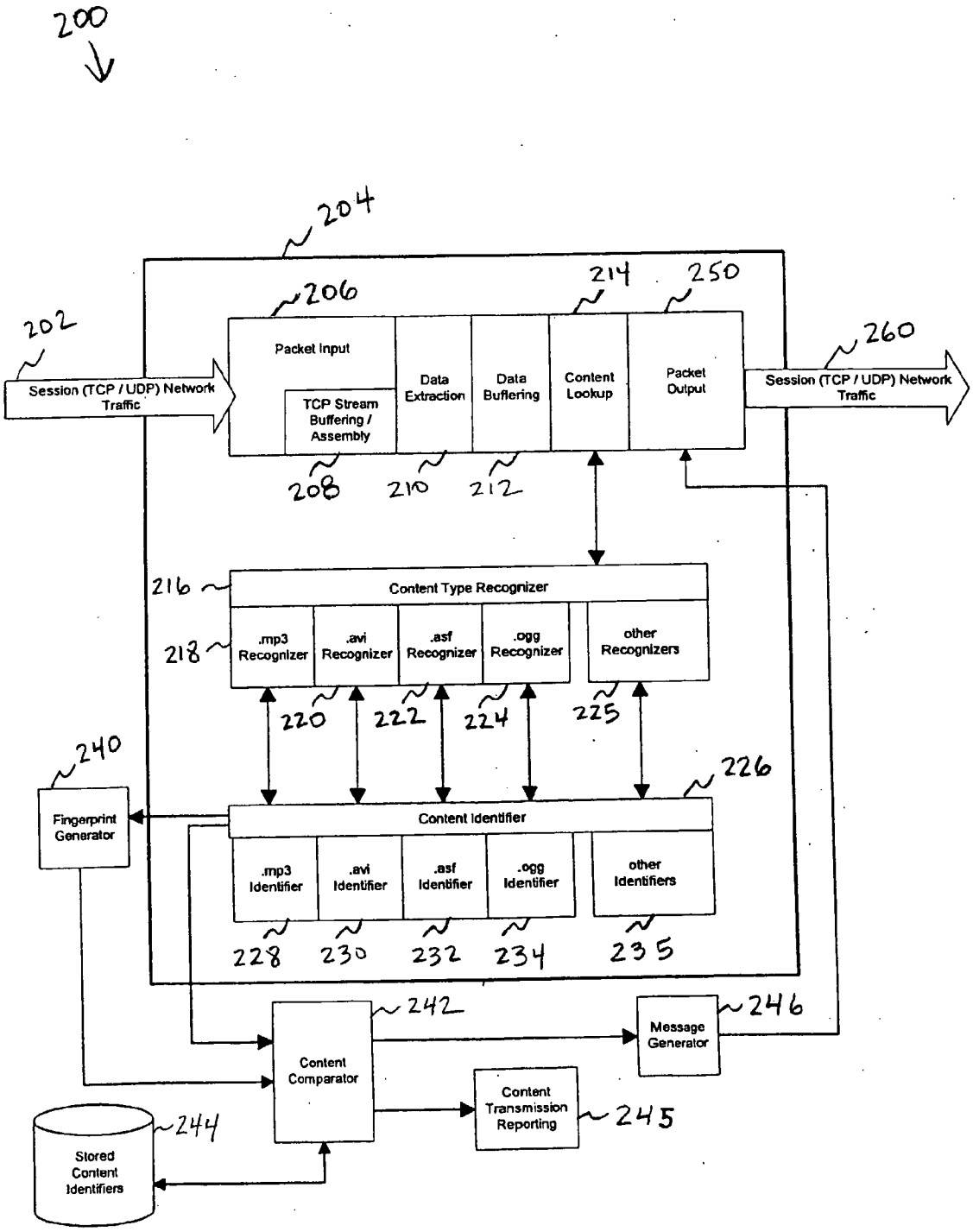

200
204
206
214
250
202
260
Session (TCP / UDP) Network Traffic
Packet Input
TCP Stream Buffering / Assembly
Data Extraction
Data Buffering
Content Lookup
Packet Output
Session (TCP / UDP) Network Traffic
208
210
212
216
Content Type Recognizer
218
.mp3 Recognizer
.avi Recognizer
.asf Recognizer
.ogg Recognizer
other Recognizers
220
222
224
225
240
226
Fingerprint Generator
Content Identifier
.mp3 Identifier
.avi Identifier
.asf Identifier
.ogg Identifier
other Identifiers
228
230
232
234
235
242
246
Content Comparator
Message Generator
Content Transmission Reporting
245
244
Stored Content Identifiers

FIG. 4

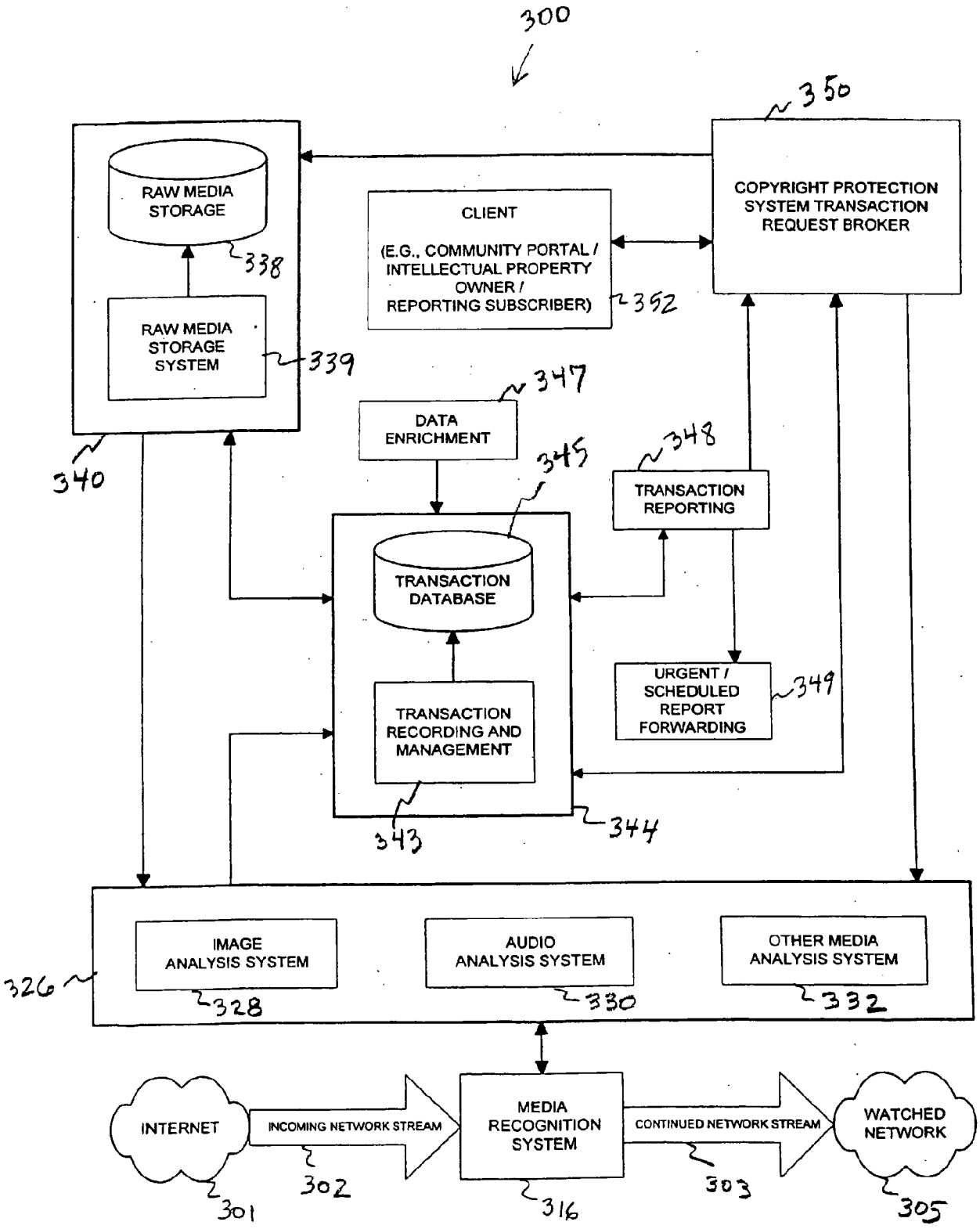

300
350
COPYRIGHT PROTECTION SYSTEM TRANSACTION REQUEST BROKER
RAW MEDIA STORAGE
338
RAW MEDIA STORAGE SYSTEM
339
340
CLIENT
(E.G., COMMUNITY PORTAL / INTELLECTUAL PROPERTY OWNER / REPORTING SUBSCRIBER)
352
347
DATA ENRICHMENT
348
TRANSACTION REPORTING
345
TRANSACTION DATABASE
TRANSACTION RECORDING AND MANAGEMENT
343
344
URGENT / SCHEDULED REPORT FORWARDING
349
326
IMAGE ANALYSIS SYSTEM
328
AUDIO ANALYSIS SYSTEM
330
OTHER MEDIA ANALYSIS SYSTEM
332
INTERNET
301
INCOMING NETWORK STREAM
302
MEDIA RECOGNITION SYSTEM
316
CONTINUED NETWORK STREAM
303
WATCHED NETWORK
305

FIG. 5

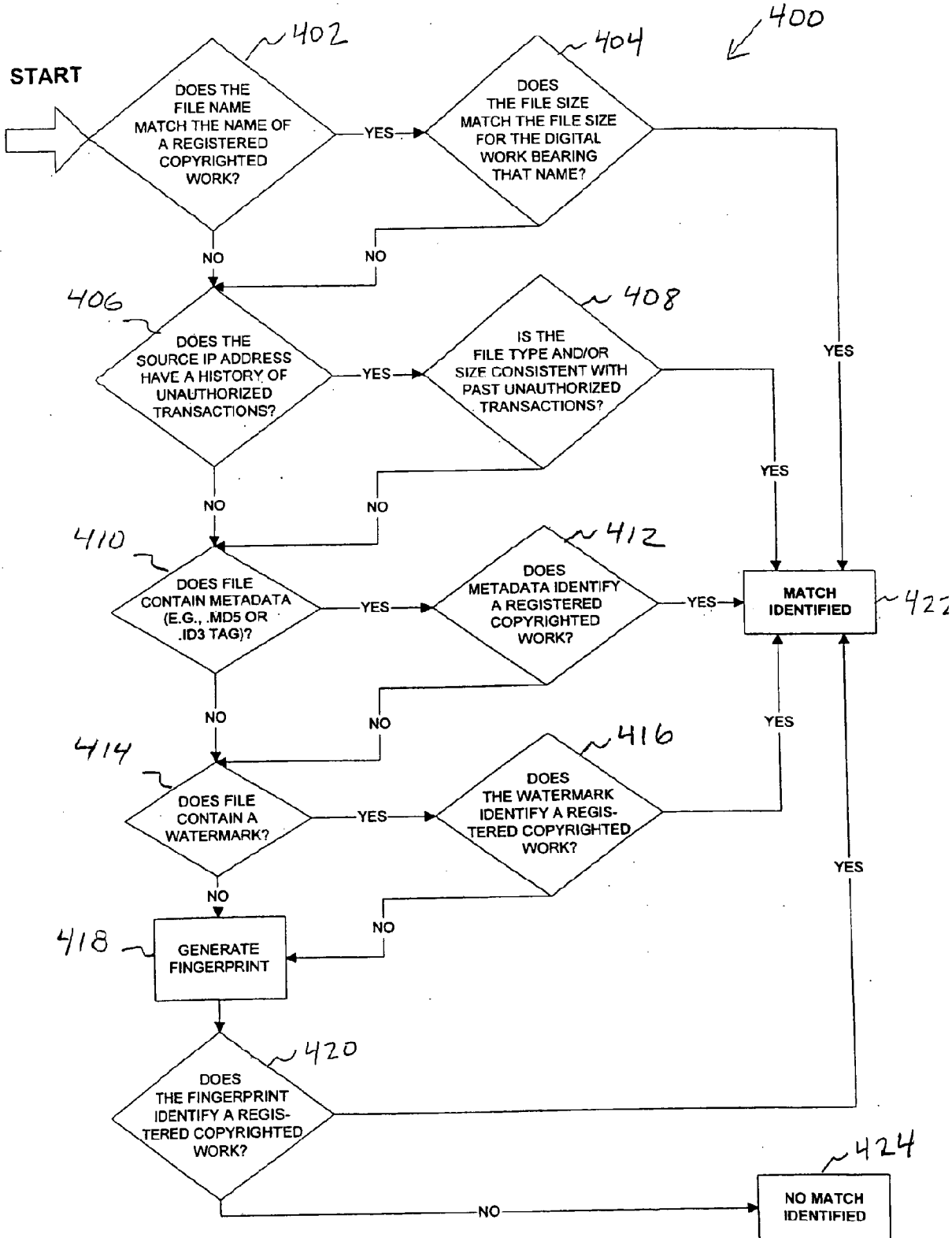

400
START
402
DOES THE FILE NAME MATCH THE NAME OF A REGISTERED COPYRIGHTED WORK?
YES
404
DOES THE FILE SIZE MATCH THE FILE SIZE FOR THE DIGITAL WORK BEARING THAT NAME?
NO
NO
406
DOES THE SOURCE IP ADDRESS HAVE A HISTORY OF UNAUTHORIZED TRANSACTIONS?
YES
408
IS THE FILE TYPE AND/OR SIZE CONSISTENT WITH PAST UNAUTHORIZED TRANSACTIONS?
YES
YES
NO
NO
410
DOES FILE CONTAIN METADATA (E.G., .MD5 OR .ID3 TAG)?
YES
412
DOES METADATA IDENTIFY A REGISTERED COPYRIGHTED WORK?
YES
MATCH IDENTIFIED
NO
NO
YES
414
DOES FILE CONTAIN A WATERMARK?
YES
416
DOES THE WATERMARK IDENTIFY A REGISTERED COPYRIGHTED WORK?
YES
NO
NO
418
GENERATE FINGERPRINT
420
DOES THE FINGERPRINT IDENTIFY A REGISTERED COPYRIGHTED WORK?
424
NO
NO MATCH IDENTIFIED

FIG. 6

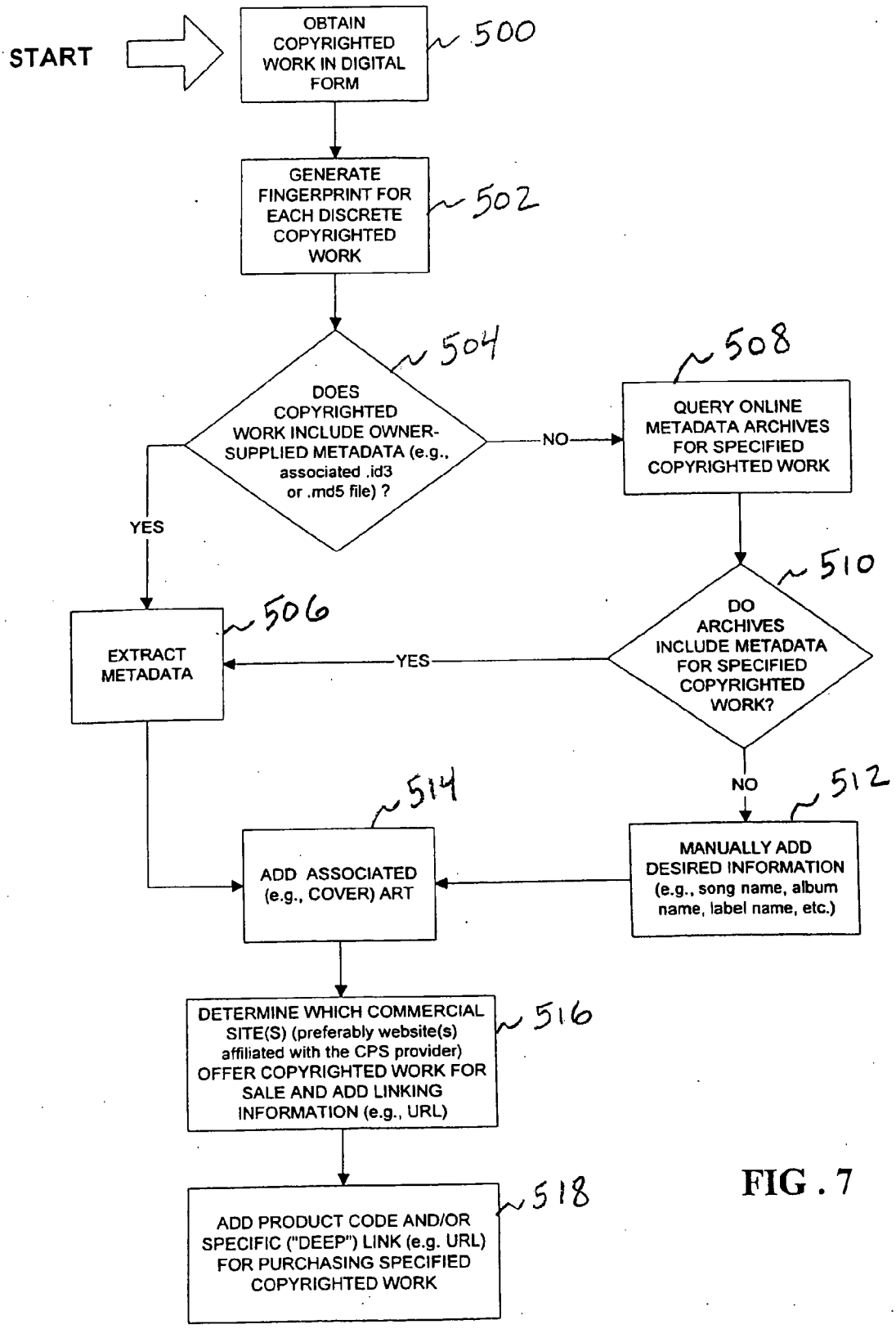

START
OBTAIN COPYRIGHTED WORK IN DIGITAL FORM
500
GENERATE FINGERPRINT FOR EACH DISCRETE COPYRIGHTED WORK
502
DOES COPYRIGHTED WORK INCLUDE OWNER-SUPPLIED METADATA (e.g., associated .id3 or .md5 file) ?
504
NO
YES
QUERY ONLINE METADATA ARCHIVES FOR SPECIFIED COPYRIGHTED WORK
508
DO ARCHIVES INCLUDE METADATA FOR SPECIFIED COPYRIGHTED WORK?
510
YES
NO
EXTRACT METADATA
506
MANUALLY ADD DESIRED INFORMATION (e.g., song name, album name, label name, etc.)
512
ADD ASSOCIATED (e.g., COVER) ART
514
DETERMINE WHICH COMMERCIAL SITE(S) (preferably website(s) affiliated with the CPS provider) OFFER COPYRIGHTED WORK FOR SALE AND ADD LINKING INFORMATION (e.g., URL)
516
ADD PRODUCT CODE AND/OR SPECIFIC ("DEEP") LINK (e.g. URL) FOR PURCHASING SPECIFIED COPYRIGHTED WORK
518

FIG. 7

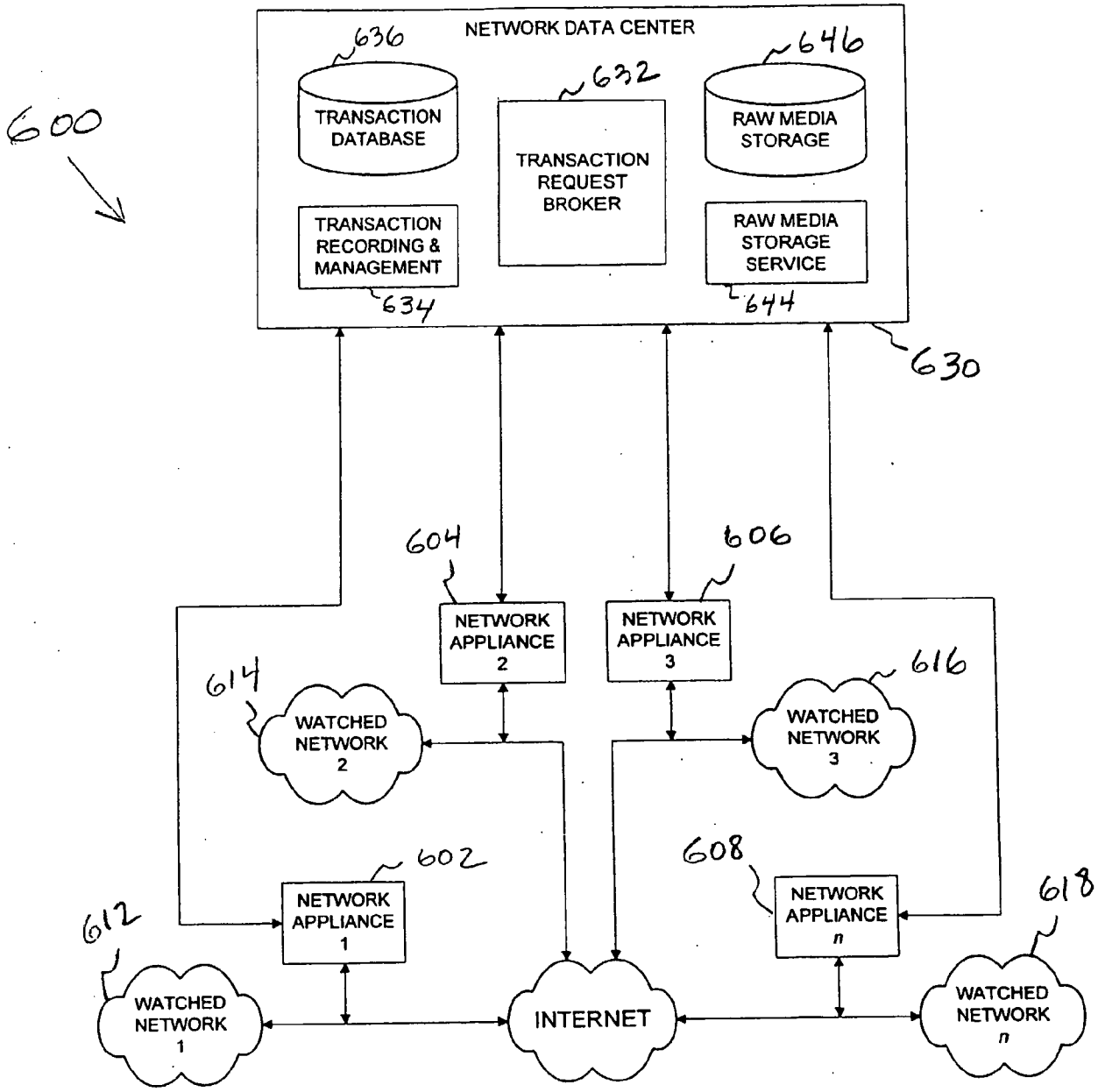

600
NETWORK DATA CENTER
636
TRANSACTION DATABASE
632
TRANSACTION REQUEST BROKER
646
RAW MEDIA STORAGE
TRANSACTION RECORDING & MANAGEMENT
634
RAW MEDIA STORAGE SERVICE
644
630
604
606
NETWORK APPLIANCE 2
NETWORK APPLIANCE 3
614
616
WATCHED NETWORK 2
WATCHED NETWORK 3
602
608
618
612
NETWORK APPLIANCE 1
NETWORK APPLIANCE n
WATCHED NETWORK 1
INTERNET
WATCHED NETWORK n

FIG. 8

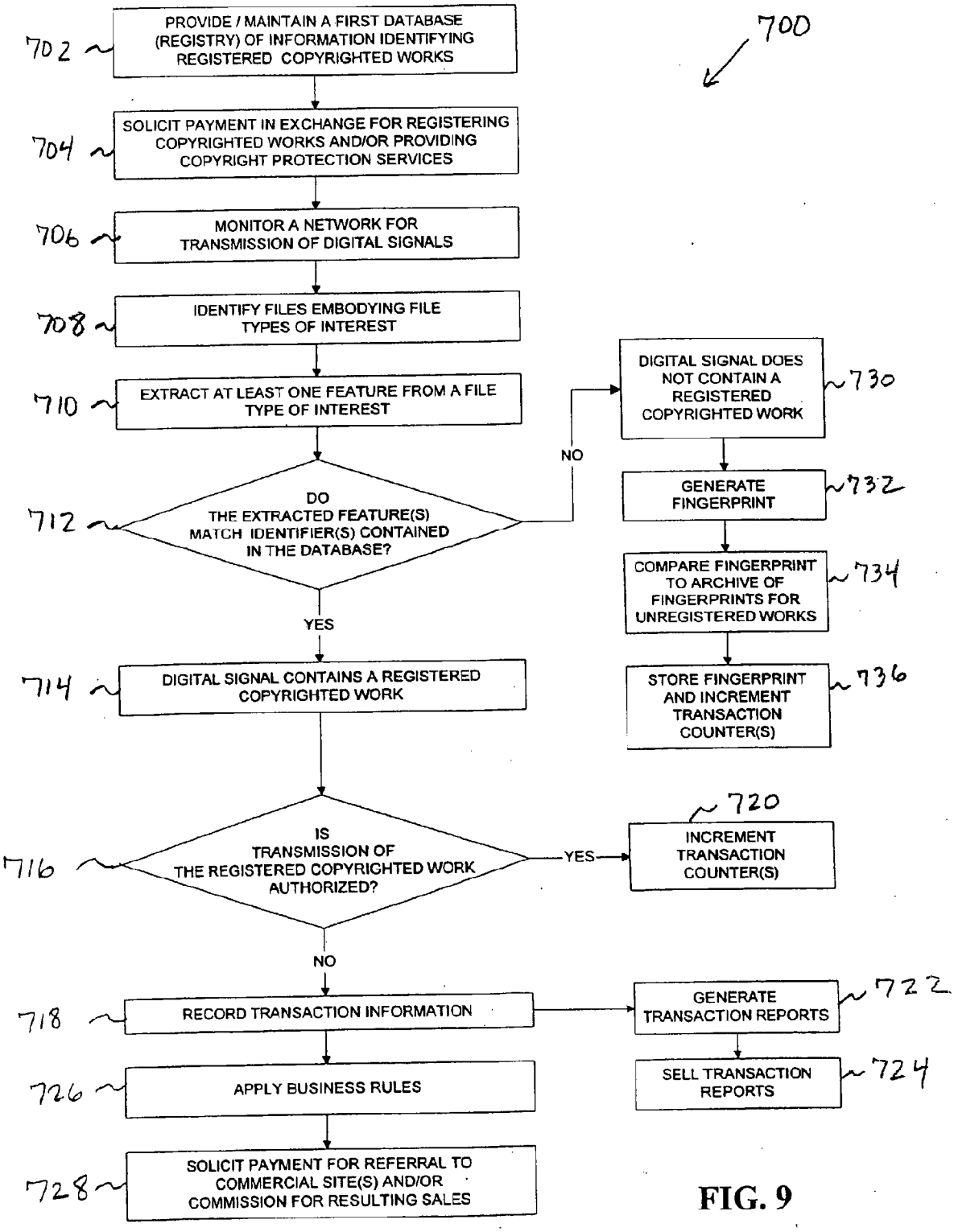

700
702
PROVIDE / MAINTAIN A FIRST DATABASE (REGISTRY) OF INFORMATION IDENTIFYING REGISTERED COPYRIGHTED WORKS
704
SOLICIT PAYMENT IN EXCHANGE FOR REGISTERING COPYRIGHTED WORKS AND/OR PROVIDING COPYRIGHT PROTECTION SERVICES
706
MONITOR A NETWORK FOR TRANSMISSION OF DIGITAL SIGNALS
708
IDENTIFY FILES EMBODYING FILE TYPES OF INTEREST
710
EXTRACT AT LEAST ONE FEATURE FROM A FILE TYPE OF INTEREST
712
DO THE EXTRACTED FEATURE(S) MATCH IDENTIFIER(S) CONTAINED IN THE DATABASE?
NO
730
DIGITAL SIGNAL DOES NOT CONTAIN A REGISTERED COPYRIGHTED WORK
732
GENERATE FINGERPRINT
734
COMPARE FINGERPRINT TO ARCHIVE OF FINGERPRINTS FOR UNREGISTERED WORKS
736
STORE FINGERPRINT AND INCREMENT TRANSACTION COUNTER(S)
YES
714
DIGITAL SIGNAL CONTAINS A REGISTERED COPYRIGHTED WORK
716
IS TRANSMISSION OF THE REGISTERED COPYRIGHTED WORK AUTHORIZED?
YES
720
INCREMENT TRANSACTION COUNTER(S)
NO
718
RECORD TRANSACTION INFORMATION
722
GENERATE TRANSACTION REPORTS
724
SELL TRANSACTION REPORTS
726
APPLY BUSINESS RULES
728
SOLICIT PAYMENT FOR REFERRAL TO COMMERCIAL SITE(S) AND/OR COMMISSION FOR RESULTING SALES

FIG. 9

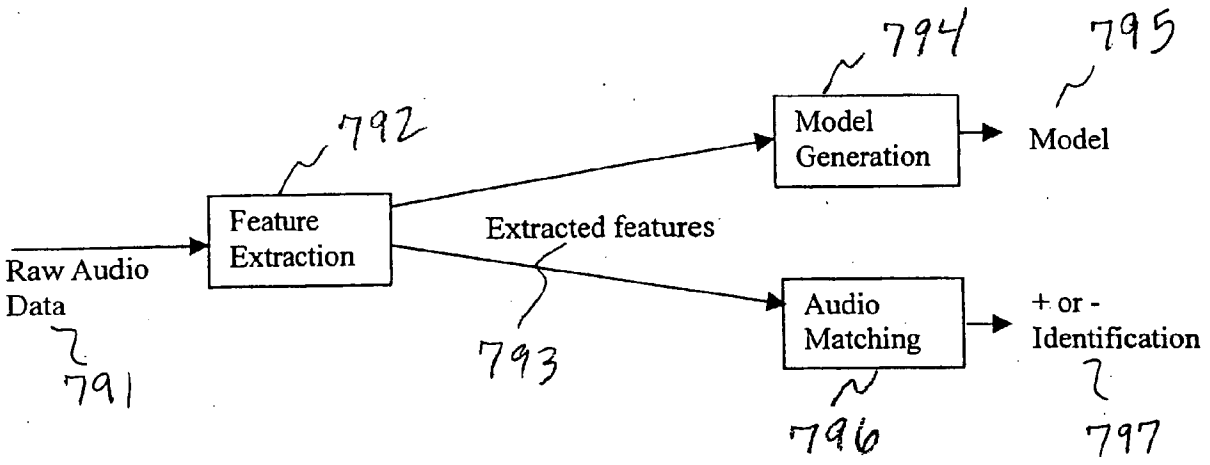

794
795
792
Model Generation
Model
Feature Extraction
Extracted features
Raw Audio Data
Audio Matching
+ or - Identification
791
793
796
797

FIG. 10

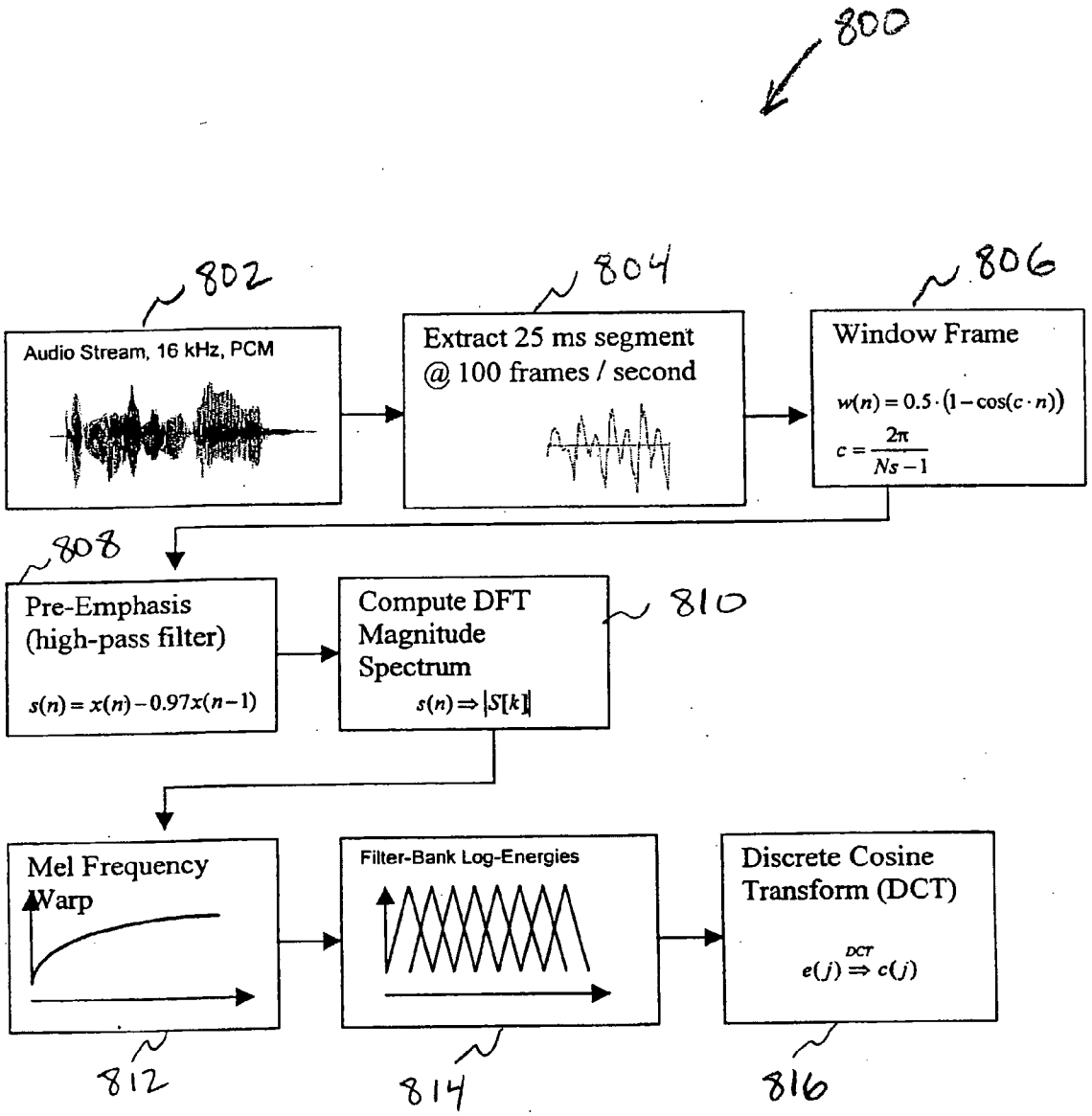

800
802
Audio Stream, 16 kHz, PCM
804
Extract 25 ms segment
@ 100 frames / second
806
Window Frame
$w(n) = 0.5 \cdot (1 - \cos(c \cdot n))$
$c = \frac{2\pi}{Ns - 1}$
808
Pre-Emphasis
(high-pass filter)
$s(n) = x(n) - 0.97x(n-1)$
Compute DFT
Magnitude
Spectrum
$s(n) \Rightarrow |S[k]|$
810
Mel Frequency
Warp
Filter-Bank Log-Energies
Discrete Cosine
Transform (DCT)
$e(j) \overset{DCT}{\Rightarrow} c(j)$
812
814
816

FIG . 11

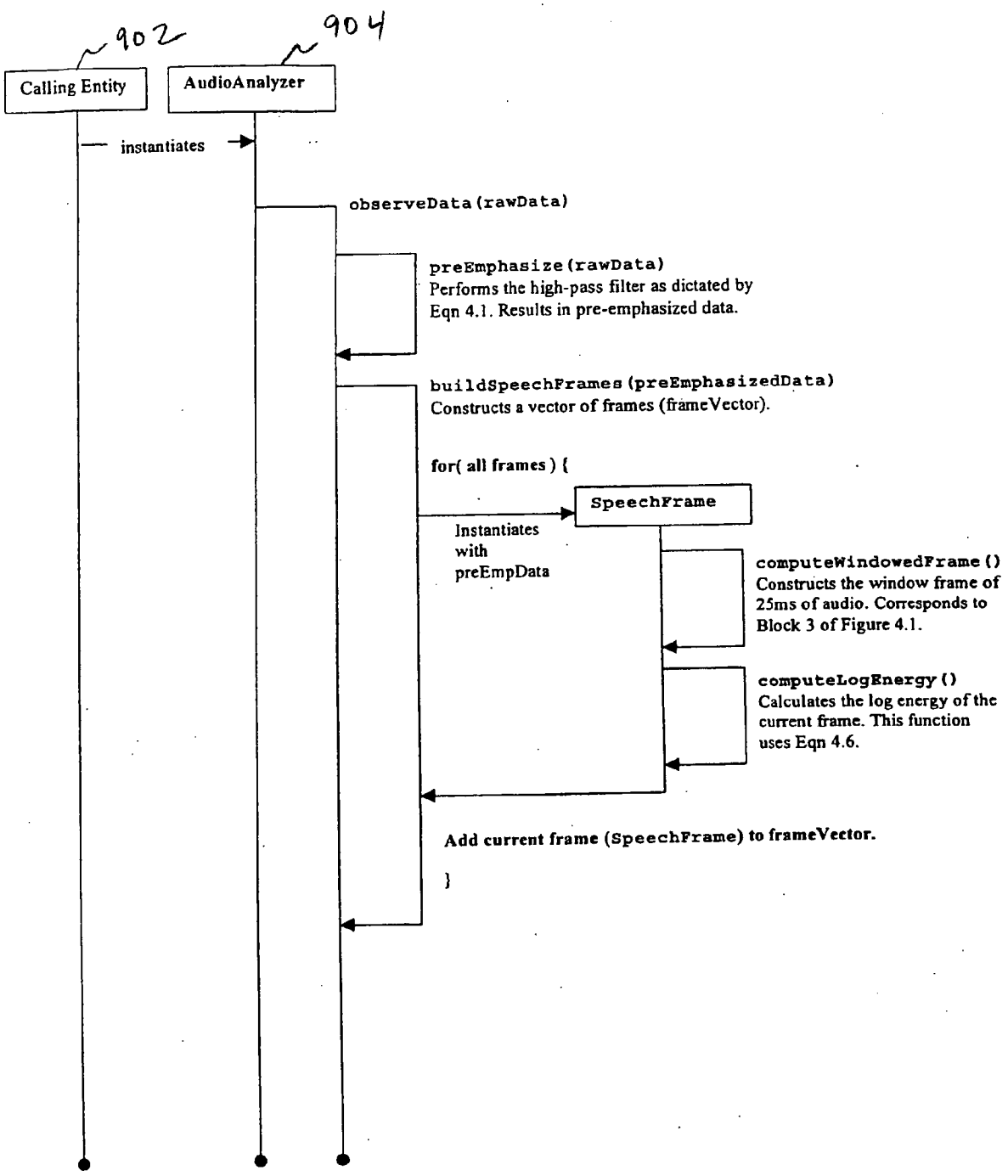

902
904
Calling Entity
AudioAnalyzer
instantiates
observeData(rawData)
preEmphasize(rawData)
Performs the high-pass filter as dictated by Eqn 4.1. Results in pre-emphasized data.
buildSpeechFrames(preEmphasizedData)
Constructs a vector of frames (frameVector).
for(all frames) {
SpeechFrame
Instantiates with preEmpData
computeWindowedFrame()
Constructs the window frame of 25ms of audio. Corresponds to Block 3 of Figure 4.1.
computeLogEnergy()
Calculates the log energy of the current frame. This function uses Eqn 4.6.
Add current frame (SpeechFrame) to frameVector.
}

FIG. 12 a

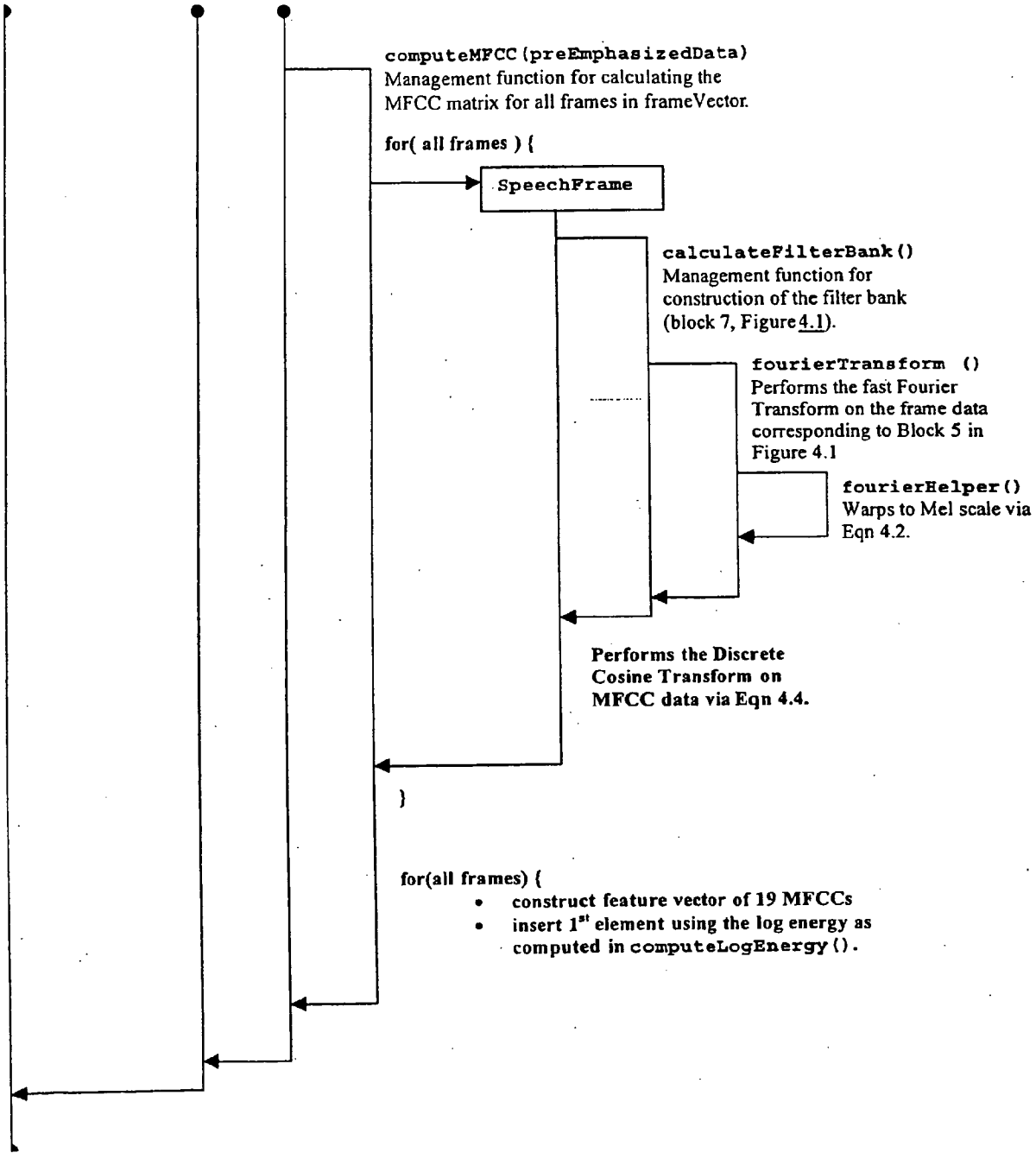

computeMFCC(preEmphasizedData)
Management function for calculating the MFCC matrix for all frames in frameVector.
for(all frames) {
SpeechFrame
calculateFilterBank()
Management function for construction of the filter bank (block 7, Figure 4.1).
fourierTransform ()
Performs the fast Fourier Transform on the frame data corresponding to Block 5 in Figure 4.1
fourierHelper()
Warps to Mel scale via Eqn 4.2.
Performs the Discrete Cosine Transform on MFCC data via Eqn 4.4.
}
for(all frames) {
construct feature vector of 19 MFCCs
insert 1st element using the log energy as computed in computeLogEnergy().

FIG. 12b

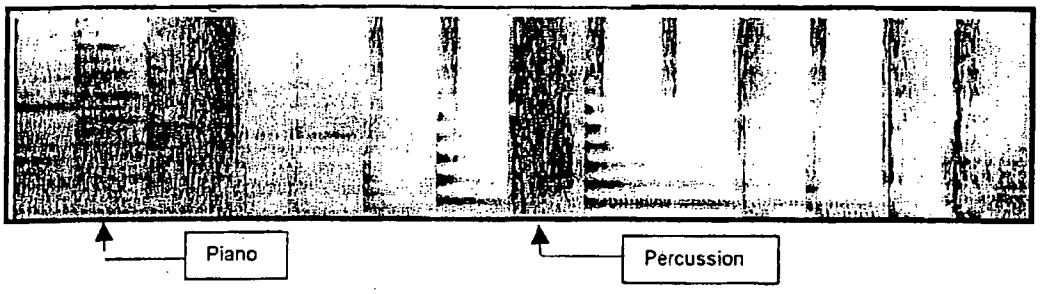


FIG . 13a

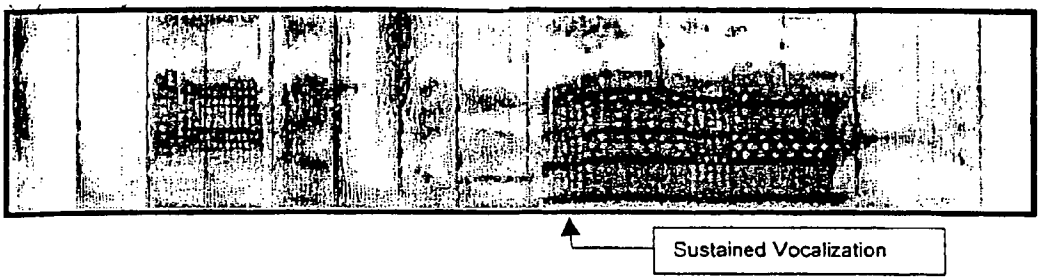


FIG. 13b

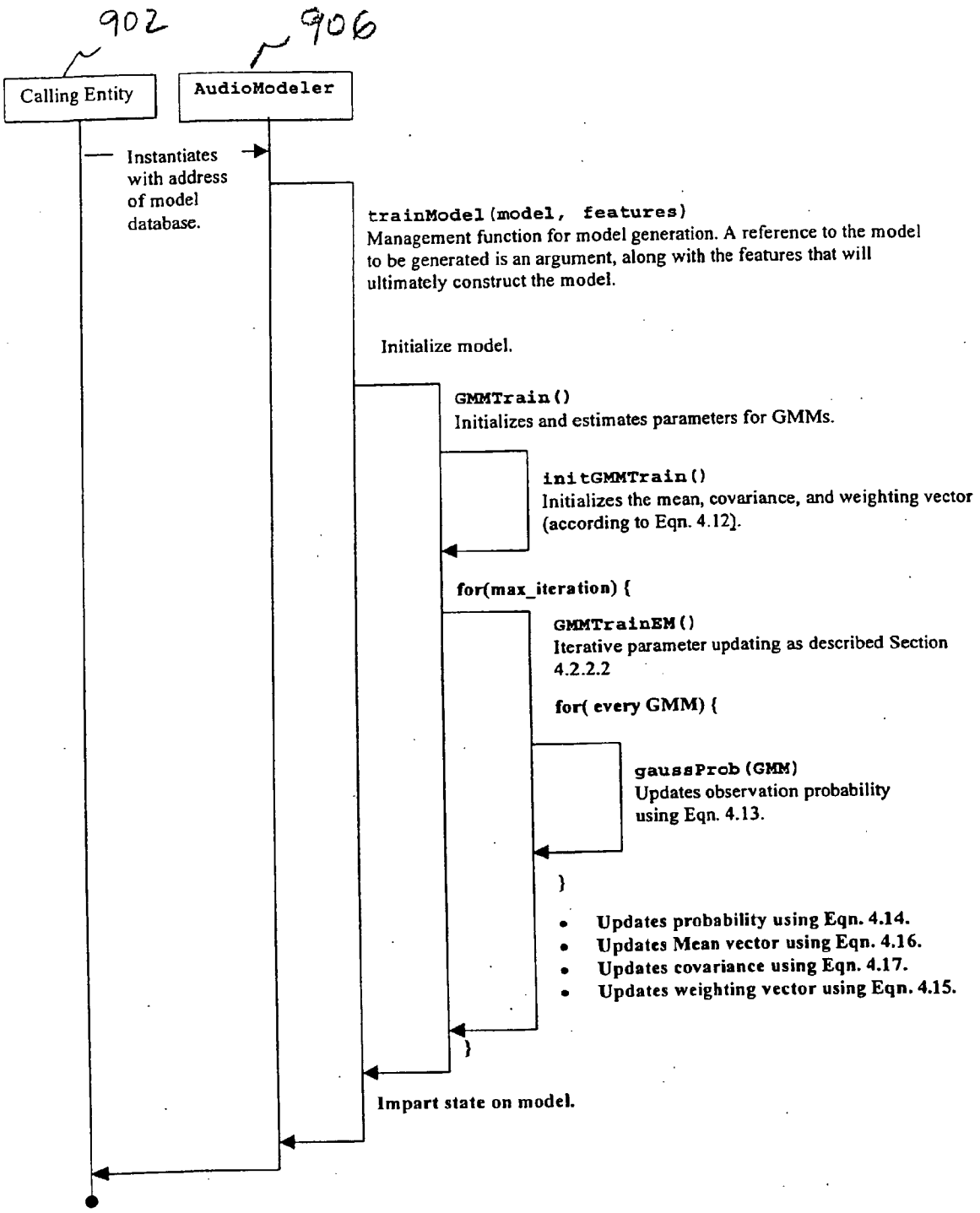

902
906
Calling Entity
AudioModeler
Instantiates with address of model database.
trainModel(model, features)
Management function for model generation. A reference to the model to be generated is an argument, along with the features that will ultimately construct the model.
Initialize model.
GMMTrain()
Initializes and estimates parameters for GMMs.
initGMMTrain()
Initializes the mean, covariance, and weighting vector (according to Eqn. 4.12).
for(max_iteration) {
GMMTrainEM()
Iterative parameter updating as described Section 4.2.2.2
for(every GMM) {
gaussProb(GMM)
Updates observation probability using Eqn. 4.13.
}
• Updates probability using Eqn. 4.14.
• Updates Mean vector using Eqn. 4.16.
• Updates covariance using Eqn. 4.17.
• Updates weighting vector using Eqn. 4.15.
}
Impart state on model.

FIG. 14

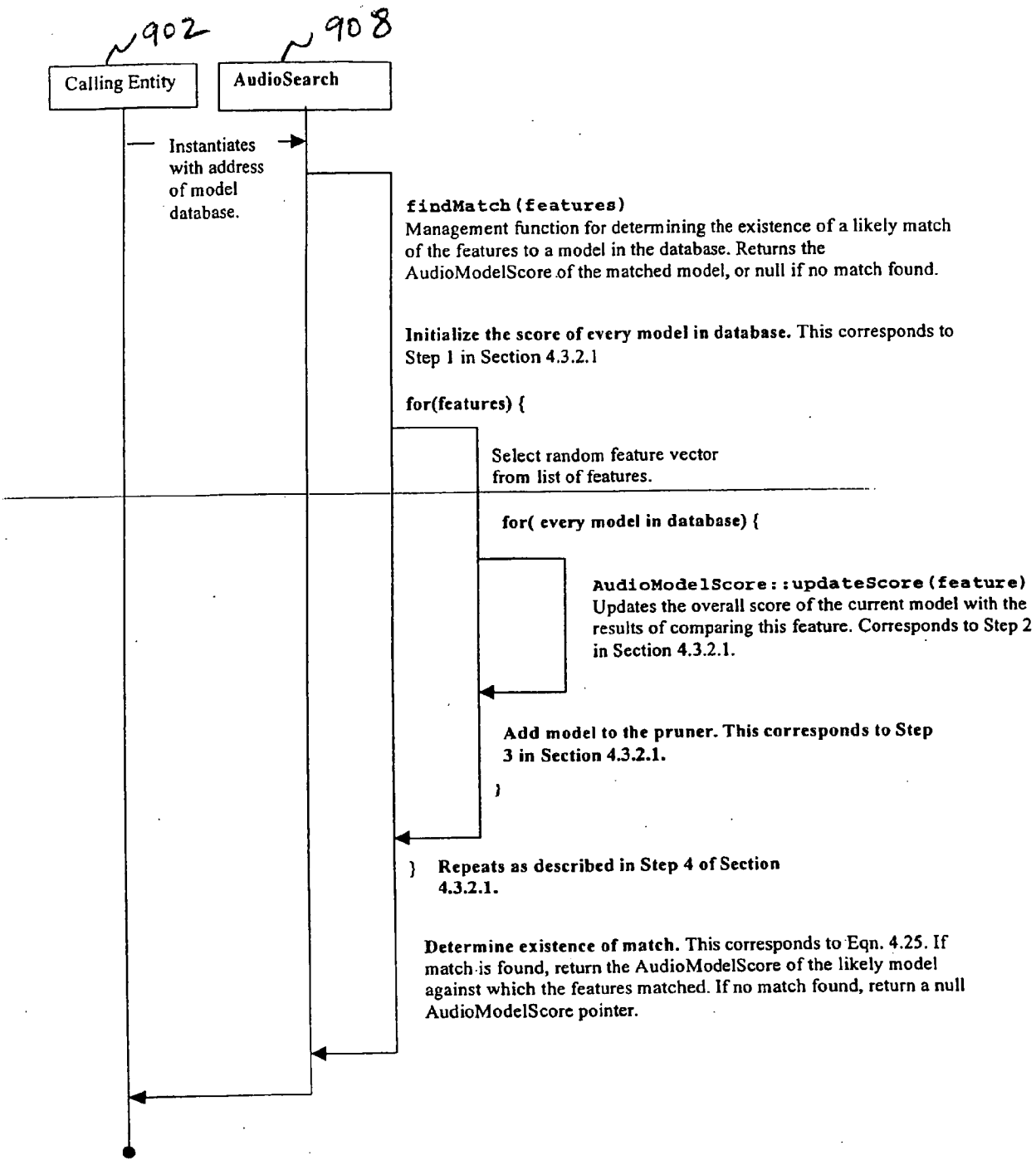

902
908
Calling Entity
AudioSearch
Instantiates with address of model database.
findMatch(features)
Management function for determining the existence of a likely match of the features to a model in the database. Returns the AudioModelScore of the matched model, or null if no match found.
Initialize the score of every model in database. This corresponds to Step 1 in Section 4.3.2.1
for(features) {
Select random feature vector from list of features.
for(every model in database) {
AudioModelScore::updateScore(feature)
Updates the overall score of the current model with the results of comparing this feature. Corresponds to Step 2 in Section 4.3.2.1.
Add model to the pruner. This corresponds to Step 3 in Section 4.3.2.1.
}
} Repeats as described in Step 4 of Section 4.3.2.1.
Determine existence of match. This corresponds to Eqn. 4.25. If match is found, return the AudioModelScore of the likely model against which the features matched. If no match found, return a null AudioModelScore pointer.

FIG . 15

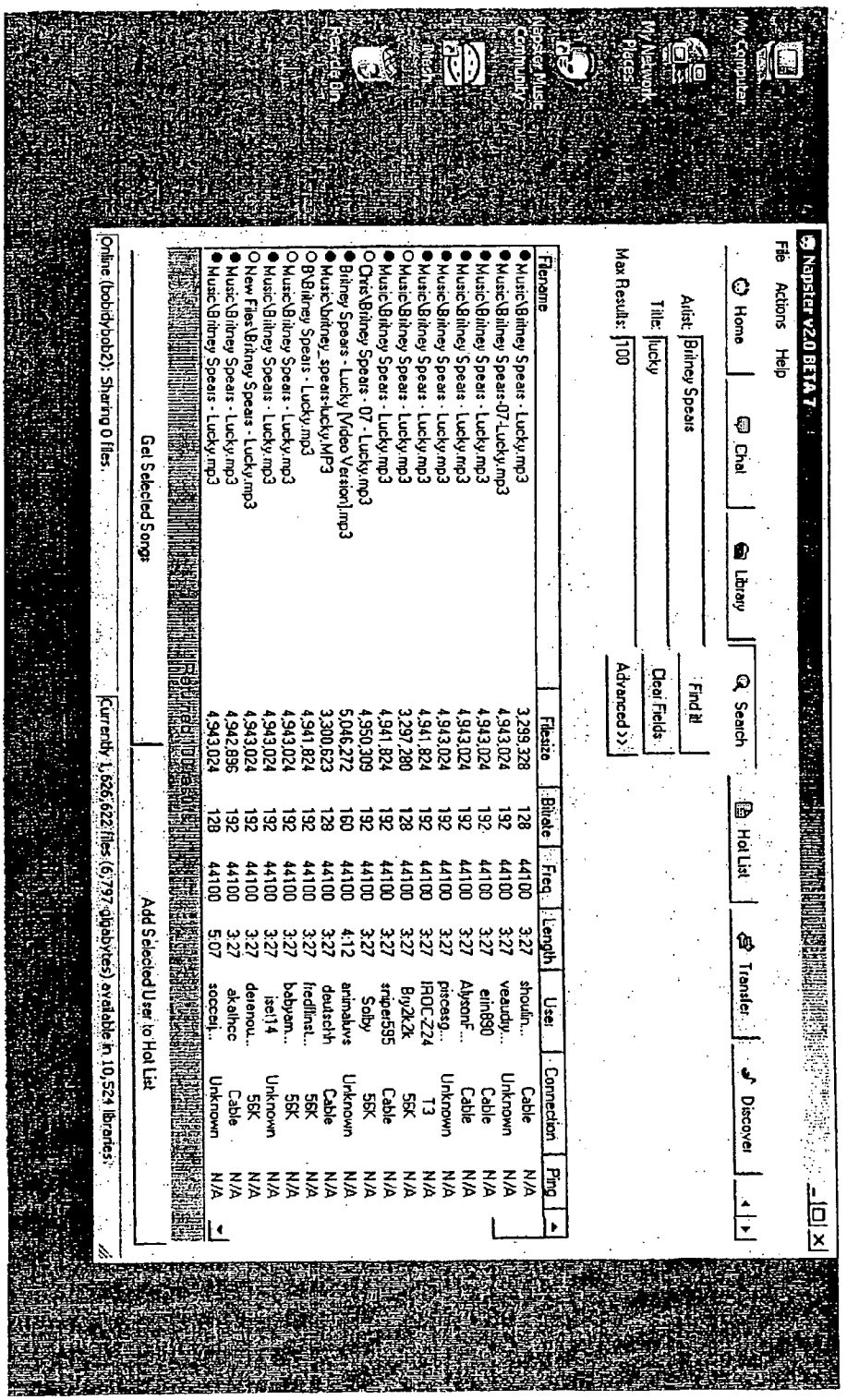

Napster v2.0 BETA 7
File Actions Help
Home
Chat
Library
Search
Hot List
Transfer
Discover
Artist: Britney Spears
Title: lucky
Max Results: 100
Find It!
Clear Fields
Advanced >>
Filename
Filesize
Bitrate
Freq
Length
User
Connection
Ping
Get Selected Songs
Add Selected User to Hot List
Online (bobidybob2): Sharing 0 files.
Currently 1,826,622 files (6,797 gigabytes) available in 10,524 libraries

FIG. 16

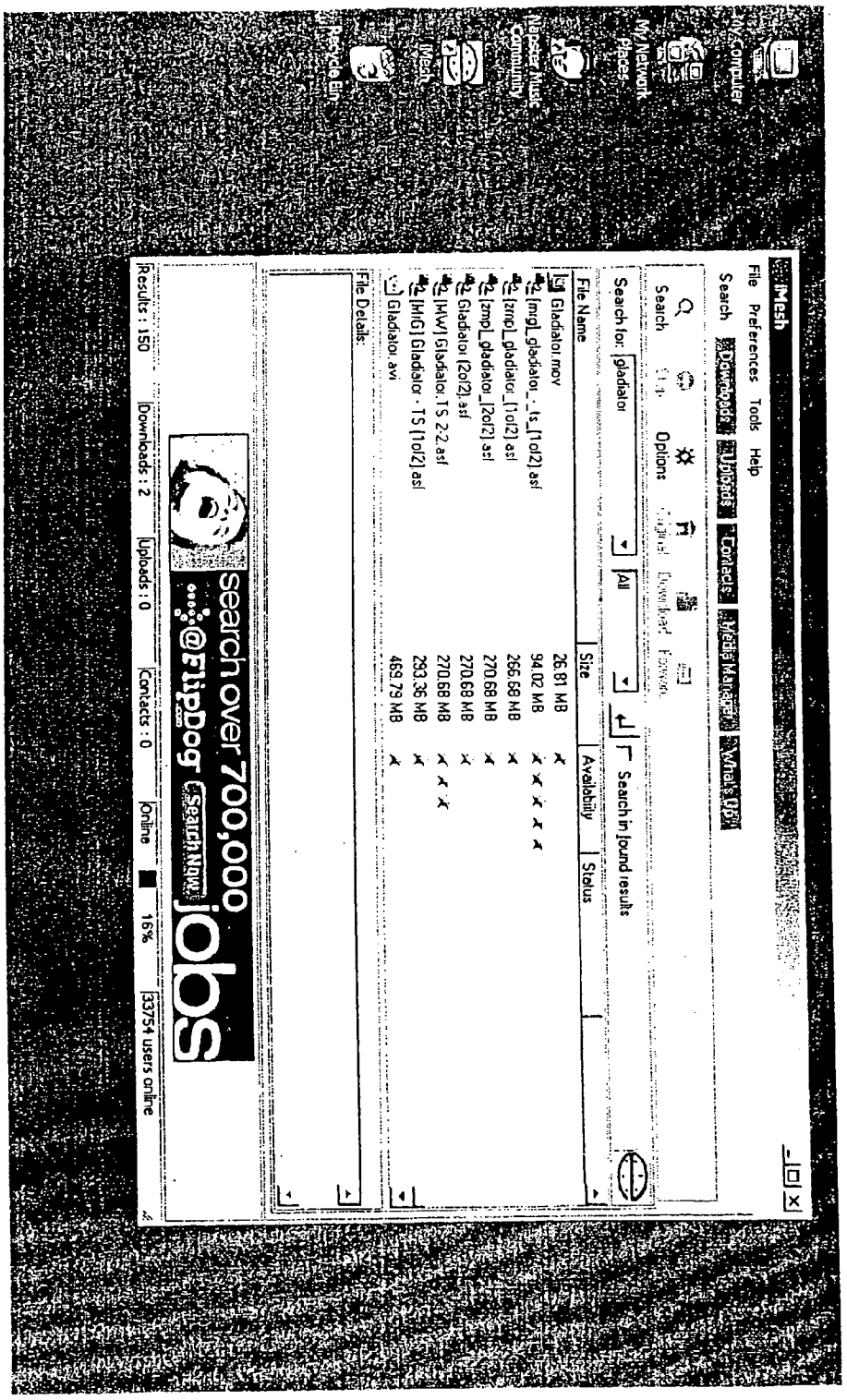

iMesh
File Preferences Tools Help
Search
Downloads
Uploads
Contacts
Media Manager
Mesh USA
Search
Options
Search for: gladiator
All
Search in found results
File Name
Size
Availability
Status
Gladiator.mov 26.81 MB
[mrg]_gladiator_-_ts_[1of2].asf 94.02 MB
[zmp]_gladiator_[1of2].asf 266.68 MB
[zmp]_gladiator_[2of2].asf 270.68 MB
Gladiator (2of2).asf 270.68 MB
[MW] Gladiator TS 2.2.asf 293.36 MB
[MG] Gladiator - TS [1of2].asf 469.79 MB
Gladiator.avi
File Details:
search over 700,000
@FlipDog Search Now! jobs
Results : 150
Downloads : 2
Uploads : 0
Contacts : 0
Online
16%
33754 users online

FIG. 17

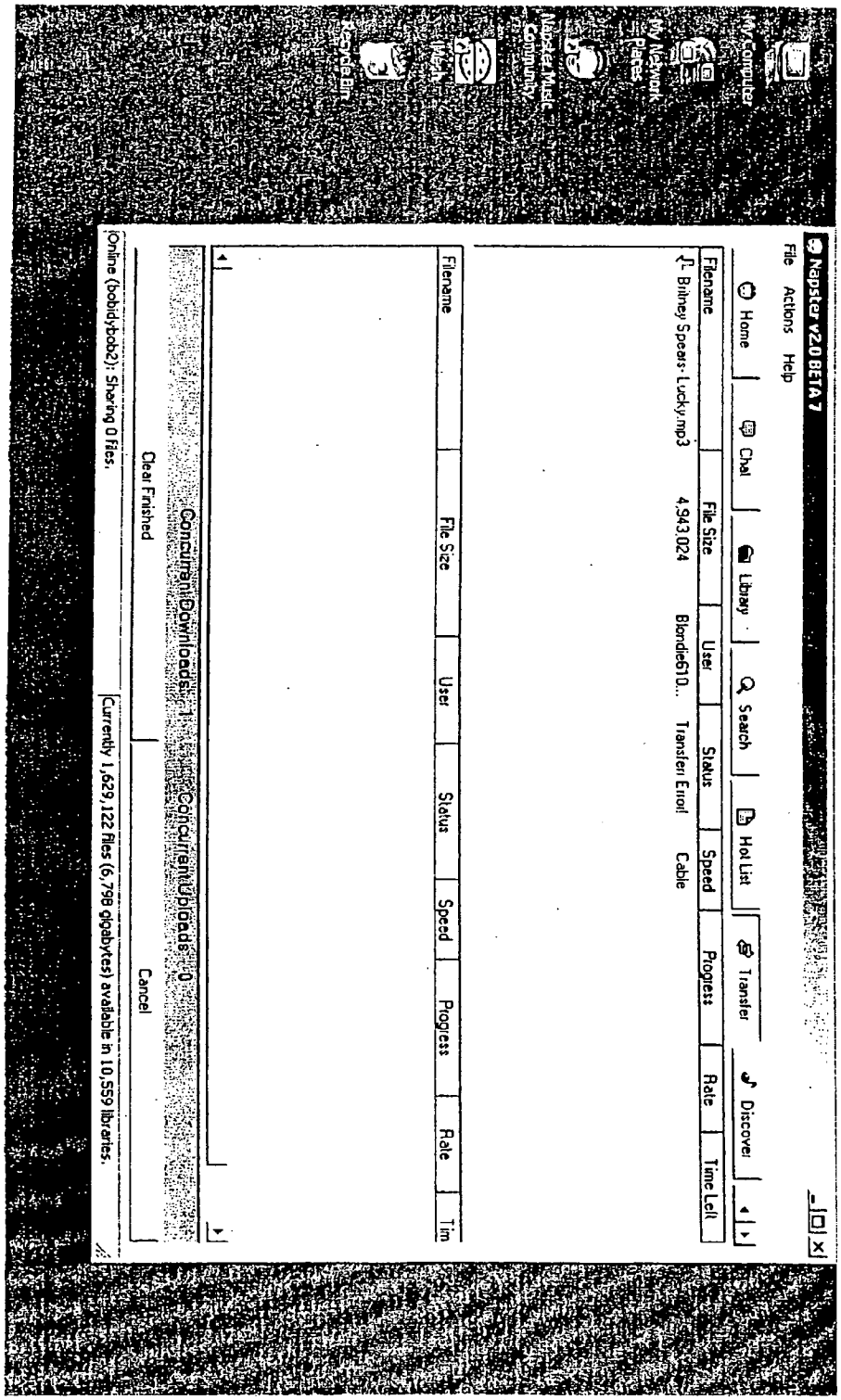

Napster v2.0 BETA 7
File Actions Help
Home
Chat
Library
Search
Hot List
Transfer
Discover
Filename
File Size
User
Status
Speed
Progress
Rate
Time Left
Britney Spears-Lucky.mp3
4,943,024
Blondie610...
Transfer Error
Cable
Concurrent Downloads: 1
Concurrent Uploads: 0
Clear Finished
Cancel
Online (bobidybob2): Sharing 0 files.
Currently 1,629,122 files (6,798 gigabytes) available in 10,559 libraries.

FIG. 18

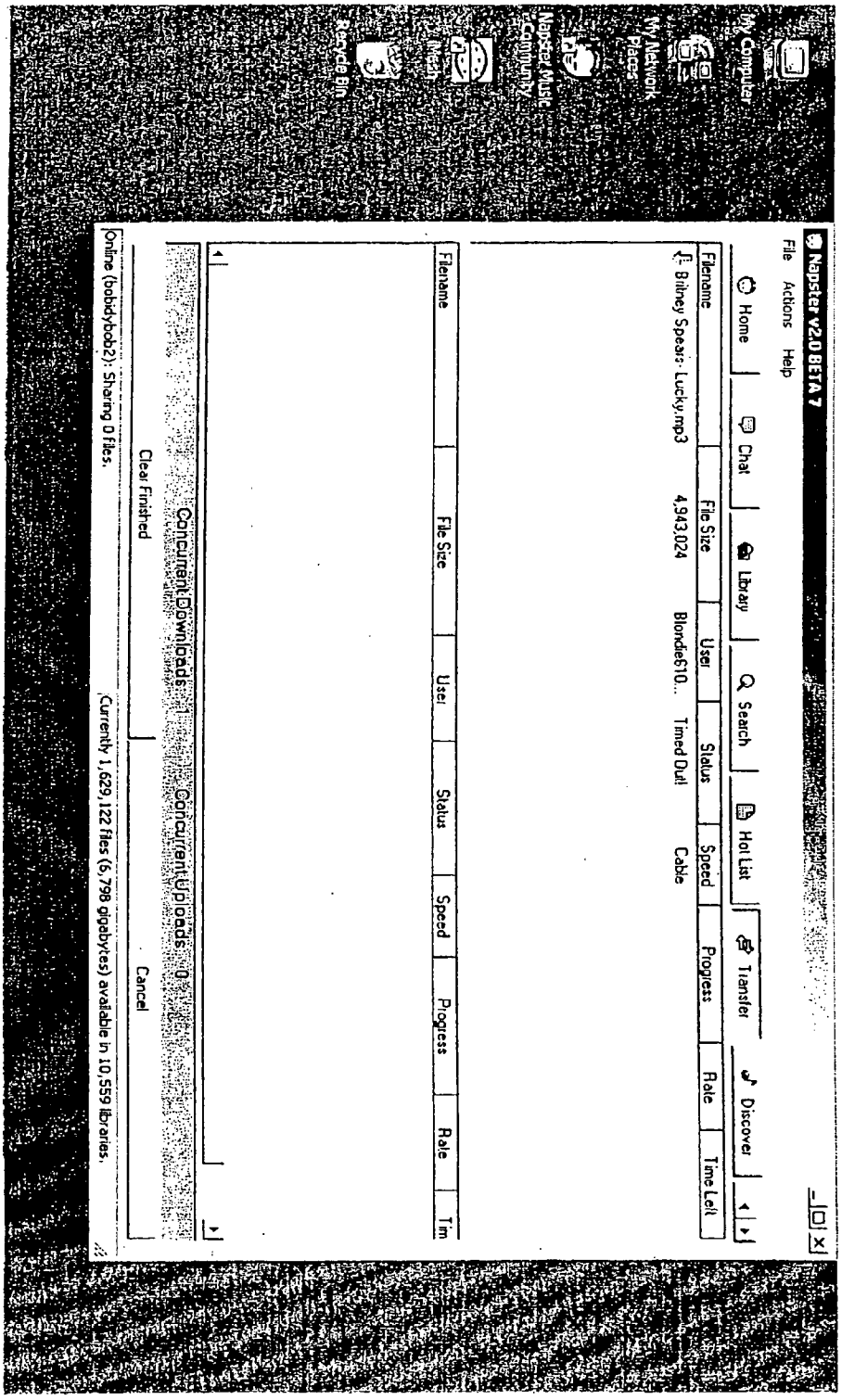

My Computer
My Network Places
Napster Music Community
Recycle Bin
Napster v2.0 BETA 7
File Actions Help
Home
Chat
Library
Search
Hot List
Transfer
Discover
Filename
File Size
User
Status
Speed
Progress
Rate
Time Left
Britney Spears: Lucky.mp3
4,943,024
Blondie610...
Timed Out!
Cable
Concurrent downloads: 0
Concurrent uploads: 0
Clear Finished
Cancel
Online (bobidybob2): Sharing 0 files.
Currently 1,629,122 files (6,798 gigabytes) available in 10,559 libraries.

FIG. 19

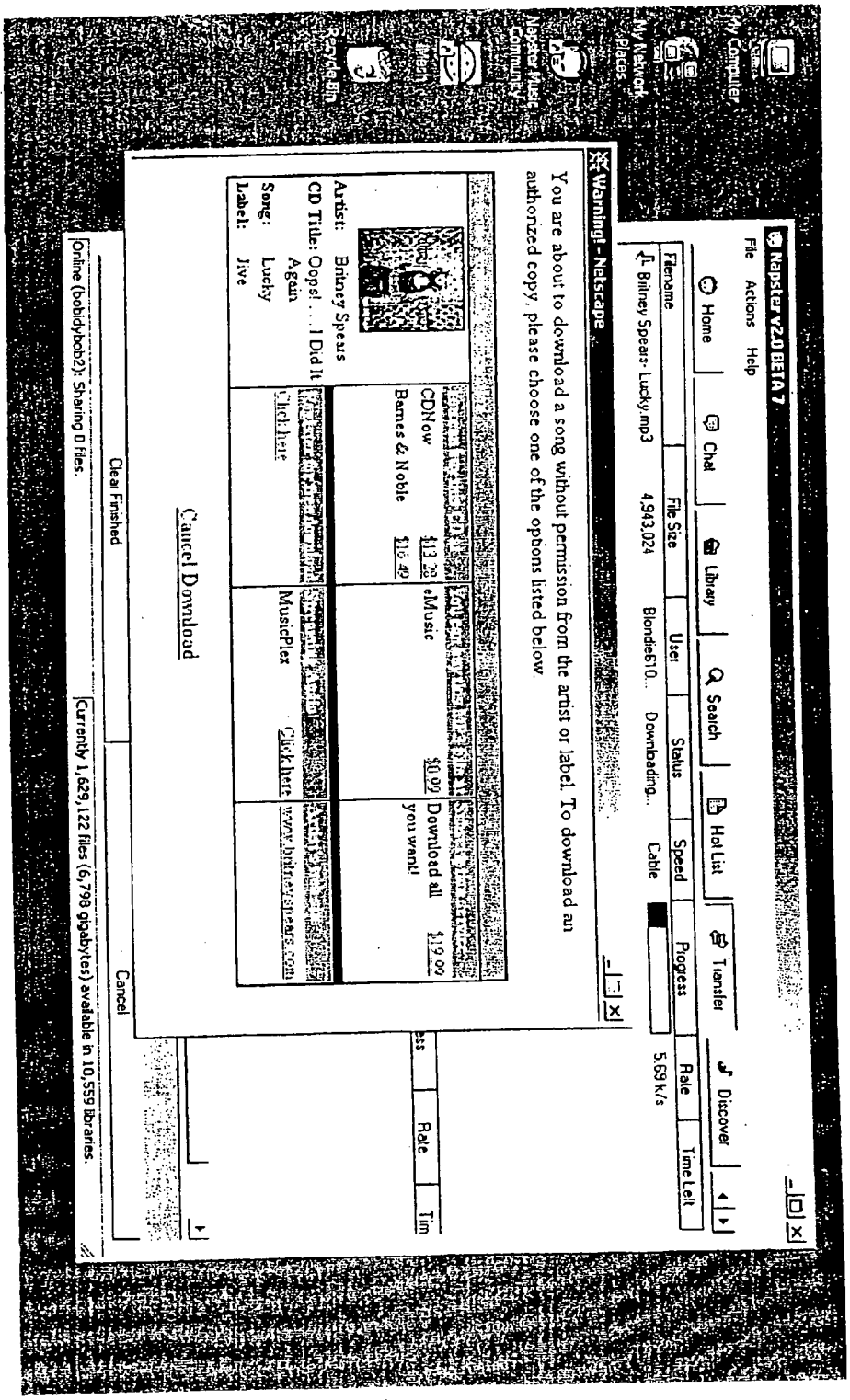

Napster v2.0 BETA 7
File Actions Help
Home
Chat
Library
Search
Hot List
Transfer
Discover
Filename
File Size
User
Status
Speed
Progress
Rate
Time Left
Britney Spears- Lucky.mp3
4,943,024
Blondie610...
Downloading...
Cable
5.69 k/s
Warning! - Netscape
You are about to download a song without permission from the artist or label. To download an authorized copy, please choose one of the options listed below.
CDNow
$13.28
eMusic
$0.99
Download all
$19.99
Barnes & Noble
$16.49
you want!
Artist: Britney Spears
CD Title: Oops! ... I Did It Again
Song: Lucky
Label: Jive
Click here
MusicPlex
Click here
www.britneyspears.com
Cancel Download
Clear Finished
Cancel
Online (bobidybob2): Sharing 0 files.
Currently 1,629,122 files (6,798 gigabytes) available in 10,559 libraries.

FIG. 20

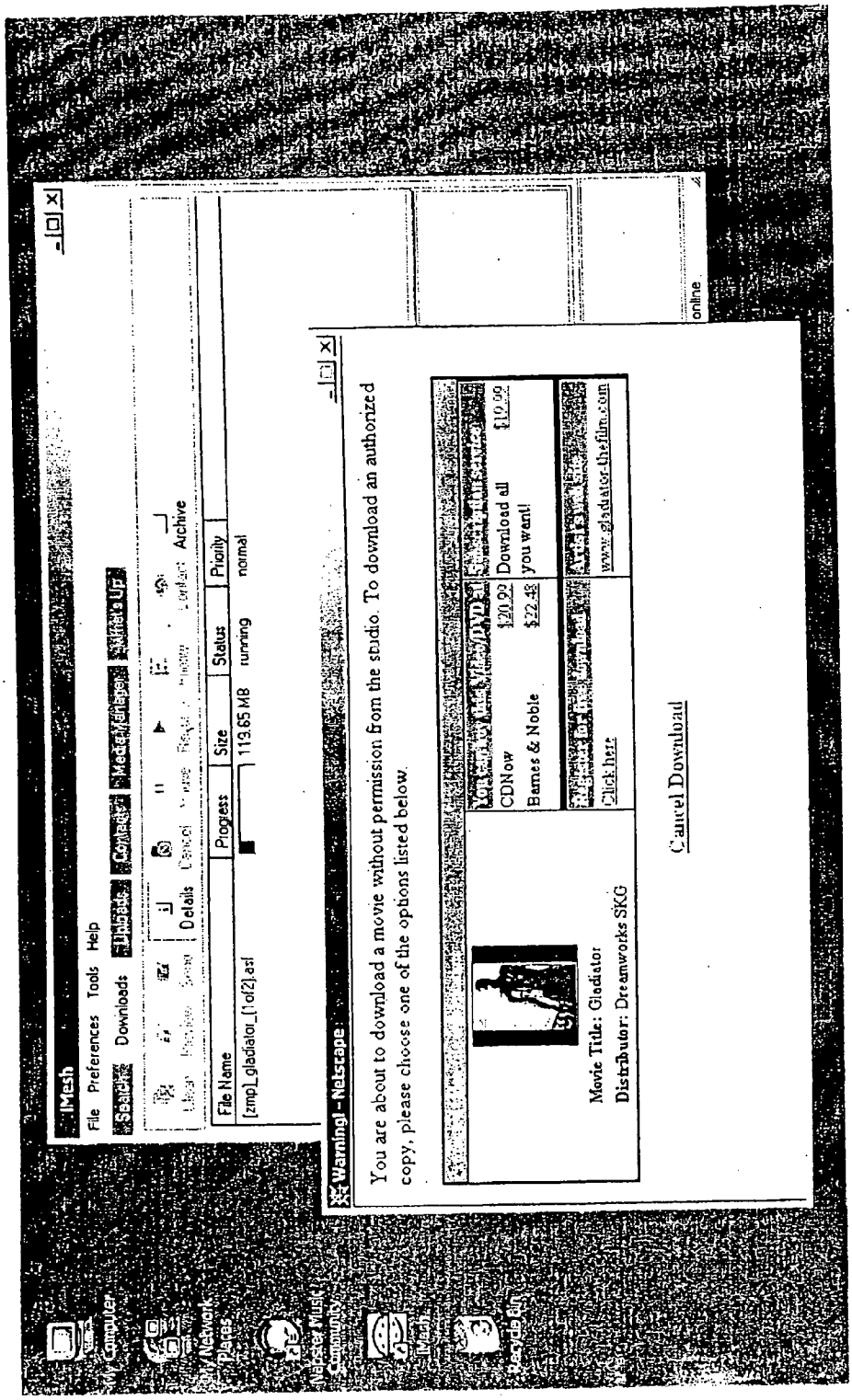

iMesh
File Preferences Tools Help
Search Downloads Friends Contacts Media Player What's Up
Details Cancel Archive
File Name Progress Size Status Priority
[zmp]_gladiator_[1of2].asf 119.65 MB running normal
Warning! - Netscape
You are about to download a movie without permission from the studio. To download an authorized copy, please choose one of the options listed below.
CDNow $20.99 Download all $10.00
Barnes & Noble $22.48 you want!
Movie Title: Gladiator
Distributor: Dreamworks SKG
Click here
www.gladiator-thefilm.com
Cancel Download
online

FIG. 21

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/10615

A. CLASSIFICATION OF SUBJECT MATTER
IPC(7) : G06F 11/30
US CL : 713/193
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 380/201,202,239,240;713/193,201;709/224; 714/38,39

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X, E --- Y, E	US 2002/0087885 A1 (PELED et al.) 04 July 2002 (04.07.2002), abstract, Fig. 9, Fig.11, pg. 4, paragraphs 96-99, pg. 5, paragraphs 101, 107-110, pg. 6, paragraphs 117,118, 1122, pg. 7, paragraphs 127-132, claims 1, 6, 10 and 13.	1,21,42,45,74,78 --- 2-19,22-41, 43-44, 46-73,79-81
Y, P	US 6,253,337 B1 (MALONEY et al.) 26 June 2001 (26.06.2001), abstract, column 2, lines 15-40, column 4, lines 15-67, column 5, lines 63-67 through column 6, lines 1-32, lines 51-59, column 7, lines 19-34column 8, lines 27-50, column 9, 14-64coulumn 10, lines 18-63.	2-19, 22-41, 43-44,46-73, 75-77, 79-81
Y, E	US 2002/0082999 A1 (LEE et al.) 27 July 2002 (27.07.2002), the enire document.	1,21,42,45,74,78
Y,P	US 6,253,193 B1 (GINTER et al.) 26 June 2001 (26.06.2001), the entire document.	2-19, 22-41, 43-44, 46-73, 75-77, 79-81
A,P	WO 02/15035 A2 (NAPSTER, INC) 21 February 2002 (21.02.2002), the entire document.	1, 21, 42, 45, 74, 78

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
14 July 2002 (14.07.2002)	07 AUG 2002
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703)305-3230	Authorized officer Gail O Hayes Telephone No. (703) 305-4274

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/10615

Continuation of B. FIELDS SEARCHED Item 3:
WEST, DIALOG, ProQuest,Dogpile. Search terms, network analysis, copy protected and surveillance, digital work and monitoring , network traffic analysis, routing analysis